doi:10.1006/jmbi.2001.4611 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **308**, 721–743

JMB

Heat-induced Unfolding of Neocarzinostatin, a Small All-β Protein Investigated by Small-angle X-ray Scattering

J. Pérez[1], P. Vachette[1], D. Russo[2], M. Desmadril[3] and D. Durand[1]*

[1]*LURE, Université Paris-Sud BP 34, 91898, Orsay Cédex France*

[2]*LLB, CE Saclay, 91191 Gif-sur-Yvette Cédex, France*

[3]*LMIP, Université Paris-Sud 91405 Orsay Cédex, France*

**Corresponding author*

Neocarzinostatin is an all-β protein, 113 amino acid residues long, with an immunoglobulin-like fold. Its thermal unfolding has been studied by small-angle X-ray scattering. Preliminary differential scanning calorimetry and fluorescence measurements suggest that the transition is not a simple, two-state transition. The apparent radius of gyration is determined using three different approaches, the validity of which is critically assessed using our experimental data as well as a simple, two-state model. Similarly, each step of data analysis is evaluated and the underlying assumptions plainly stated. The existence of at least one intermediate state is formally demonstrated by a singular value decomposition of the set of scattering patterns. We assume that the pattern of the solution before the onset of the transition is that of the native protein, and that of the solution at the highest temperature is that of the completely unfolded protein. Given these, actually not very restrictive, boundary constraints, a least-squares procedure yields a scattering pattern of the intermediate state. However, this solution is not unique: a whole class of possible solutions is derived by adding to the previous linear combination of the native and completely unfolded states. Varying the initial conditions of the least-squares calculation leads to very similar solutions. Whatever member of the class is considered, the conformation of this intermediate state appears to be weakly structured, probably less than the transition state should be according to some proposals. Finally, we tried and used the classical model of three thermodynamically well-defined states to account for our data. The failure of the simple thermodynamic model suggests that there is more than the single intermediate structure required by singular value decomposition analysis. Formally, there could be several discrete intermediate species at equilibrium, or an ensemble of conformations differently populated according to the temperature. In the latter case, a third state would be a weighted average of all non native and not completely unfolded states of the protein but, since the weights change with temperature, no meaningful curve is likely to be derived by a global analysis using the simple model of three thermodynamically well-defined states.

Keywords: protein unfolding; small-angle scattering; intermediate; ensemble of conformations; β-sheet protein

Introduction

Understanding the folding mechanism of β-sheet proteins has been a matter of considerable interest in the last few years. In their recent review, Capaldi & Radford[1] predicted that new insights would rapidly arise from experiments to come in the following years. Indeed, a series of experimental works has since been published, focussing mainly

Abbreviations used: Ig, immunoglobulin; NCS, neocarzinostatin; DSC, differential scanning calorimetry; SAXS, small-angle X-ray scattering; SVD, singular value decomposition; Gdn, guanidinium.

E-mail address of the corresponding author: durand@lure.u-psud.fr

0022-2836/01/040721–23 $35.00/0

on the detection and indirect characterisation of folding kinetic transition states.[2–8] Particularly noteworthy is the work by Clarke and co-workers, who have initiated a systematic experimental study of the folding mechanism of the immunoglobulin-like (Ig-like) fold, a fold common to several protein families that do not necessarily share evolutionary links.[9–14] The authors put forward the idea that a folding motif that is conserved without sequence homology must exist because of intrinsic thermodynamic properties, which they want to elucidate. In that series of papers, chemically induced unfolding and refolding kinetics are monitored using stopped-flow fluorescence. In particular, the existence of a clear correlation between folding rate and stability among six wild-type proteins possessing the Ig-like fold has been shown, suggesting the existence of a common folding pathway.[13] It was noted that, despite these common features, only the three less stable proteins fold with a two-state mechanism, whilst kinetic intermediates have been shown indirectly to exist for the three most stable proteins.

We have undertaken the study of the folding of an all-β protein, neocarzinostatin (NCS), structured by an Ig-like seven-stranded Greek key β-barrel,[15,16] which belongs to a family of chromophore-carrying anti-tumour proteins of bacterial origin. NCS is a 113 amino acid residue long protein[17] containing a tightly but noncovalently bound chromophore.[18] The isolated chromophore exhibits the full biological activity of NCS. It binds to DNA with high affinity and produces DNA damage by radical reactions.[19] The protein component, apo-neocarzinostatin, which is the object of the present work, serves primarily as a carrier protecting the chromophore from hydrolysis and has no cytotoxic activity.[20] NCS possesses two short disulphide bridges, each located within a different loop, the first in the cleft close to the chromophore-binding site (C37-C47), the second at the bottom of the β-barrel (C88-C93).

As a first step in the analysis of the folding pathways of NCS, we performed equilibrium unfolding experiments, both thermally and chemically induced. Chemically induced unfolding was essentially monitored by differential scanning calorimetry (DSC); the results are presented elsewhere.[21] The heat-induced unfolding of NCS was first followed by DSC and fluorescence. The results suggest the existence of one or several transition intermediates that could not be detected directly by either of these two techniques. Moreover, the stability of NCS, as measured by Trp fluorescence, $\Delta G_{D-N} = 7.3(\pm 0.1)$ kcal mol^{-1} (1 kcal = 4.184 J) (Nicaise, Valerio and M.D., unpublished results), lies in the range of values for which the existence of a kinetic folding intermediate was demonstrated by Clarke and colleagues.[13] Beyond the mere detection of intermediates, we aimed at reaching some kind of structural characterisation of these species, and for this turned to small-angle X-ray scattering (SAXS), a technique used to investigate structure in solution. The small-angle scattering method has already been put to use to detect and characterise (un)folding intermediates both at equilibrium[22–33] and kinetically, using fast mixing techniques.[34–40] As a first step in this study, we have undertaken an equilibrium study of the thermal unfolding of NCS.

A series of scattering curves of a solution of NCS, $I(Q,T_k)$, where the momentum transfer Q is related to the scattering angle 2θ and to the X-ray wavelength λ by $Q = (4\pi \sin\theta)/\lambda$, were recorded at various temperature values T_k and analysed using the singular value decomposition (SVD) approach. This approach is now commonly used in the analysis of (un)folding transitions followed by SAXS.[41–46] It allows the determination of the minimum number of structural states with linearly independent scattering curves present during the unfolding transition, with no prior knowledge of the characteristics of these states. The analysis gives clear-cut evidence for the presence of at least one structural intermediate between the native state and the unfolded state of NCS. Scattering curves of the intermediate state compatible with the data can be determined by a least-squares minimisation and are characteristic of a weakly structured state. This set of patterns yields values of the radius of gyration between 21.3 Å and 26 Å, compared with 14 Å and 26 Å, the respective values for the native and unfolded states. In the following presentation, a strong emphasis is put on a critical evaluation of the way to extract accurate information from SAXS studies of protein (un)folding.

Results

Differential scanning calorimetry and fluorescence

Calorimetric measurements on a solution of NCS in a 60 mM phosphate buffer (pH 7.0) were performed to determine the protein unfolding temperature. The heating rate was 1 K min^{-1} and the protein concentration, 3 mg ml^{-1}, was as close as possible to that used in the SAXS experiments. The evolution of the NCS specific capacity as a function of temperature is shown in Figure 1(a). A single symmetric peak, located at about 68 °C, is observed. Above 80 °C, the constancy of the specific capacity is indicative of a maximally unfolded state of the protein.

The experimental curve was first analysed with a strict two-state model of transition (calorimetric enthalpy ΔH_{cal} identical with van't Hoff ΔH_{vH}) (Figure 1(a)), then with a model that allows ΔH_{cal} and ΔH_{vH} to differ, accounting for a possible non-uniqueness of the second state (Figure 1(b)). In both cases, the fitting procedure yields a value t_m equal to 68.6 °C for the mid-point transition temperature. A value of $\Delta H_{cal} = 104$ kcal mol^{-1} is derived for the enthalpy by using the first model, whereas the second model yields a significantly

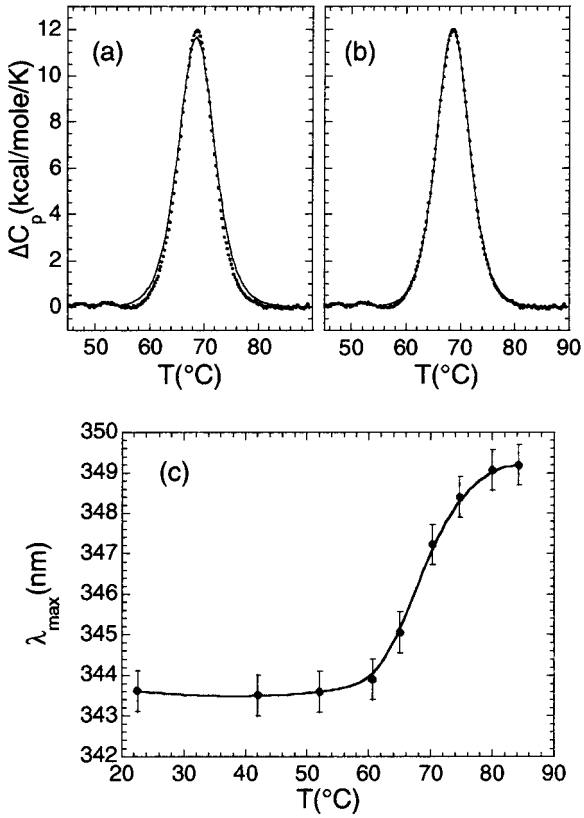

Figure 1. Differential scanning calorimetry and fluorescence monitoring of NCS unfolding. (a) and (b) Temperature-dependence of NCS specific molar heat capacity. Dots, experimental data; continuous line (a) fit by a strict two-state model $\Delta H_{cal} = \Delta H_{vH} = 104$ kcal mol^{-1}, (b) fit allowing $\Delta H_{cal} \neq \Delta H_{vH}$, $\Delta H_{cal} = 97.3$ kcal mol^{-1} and $\Delta H_{vH} = 115$ kcal mol^{-1}. (c) Temperature-dependence of the wavelength corresponding to the maximum fluorescence intensity of Trp39 and Trp83. Dots, experimental data; continuous line, fit by a strict two-state model.

better fit to the experimental curve with $\Delta H_{cal} = 97.3$ kcal mol^{-1} and $\Delta H_{vH} = 115$ kcal mol^{-1}. These results are independent of the heating rate, at least between 0.5 K min^{-1} and 2 K min^{-1}, and of the protein concentration between 0.6-3 mg ml^{-1}. Most of the NCS molecules fold back to the native state after cooling, as shown by performing a second temperature scan on the same sample. The curves obtained during the two successive heating processes present exactly the same shape and the same transition temperature. Only a 5% decrease of the peak intensity is detected in the second curve, indicating that a small amount of protein could not refold, at least not completely, during the cooling step. No aggregation is observed, which would appear as an exothermic process, making the transition peak strongly asymmetric.

The unfolding transition was also monitored using the emission spectrum of tryptophan fluorescence. The curve shown in Figure 1(c) was obtained by plotting the wavelength of the maximum of the spectrum as a function of temperature. Again, the transition appears to be co-operative and to be complete around 80 °C. A first approximation thermodynamic analysis of this curve, using a strict two-state transition model, yields $t_m = 68$ °C and $\Delta H = 66.5$ kcal mol^{-1}. Although the fluorescence data do not by themselves allow more than two-states to be distinguished, the marked difference between the values of ΔH obtained by this technique and by calorimetry suggests that the unfolding transition might not be a simple transition between two well-defined states.

SAXS

The folded protein

Before analysing the SAXS patterns of the native protein, it has to be checked that coulombic repulsions between identical proteins are effectively screened by the ionic strength of the 60 mM phosphate buffer in such a way that the interaction effect on the scattering curves is negligible. To this aim, SAXS patterns of the folded protein were recorded as a function of the protein concentration c, and at two temperatures below the unfolding transition, 22 °C and 53 °C (data not shown). The values of the radius of gyration $R_g(c)$ and of the intensity at zero angle $I(0,c)$ were derived from a Guinier analysis. The R_g value at 2.5 mg ml^{-1} was found to be 14(±0.2) Å at both temperatures. R_g values showed a very weak decrease with concentration, lower than 0.3 Å and 0.5 Å at 22 °C and 53 °C respectively in the explored range (2.5 mg ml^{-1} < c < 10 mg ml^{-1}) The virial coefficient A_2 was then estimated after plotting $c/I(0,c)$ as a function of c (equation (11) in Materials and Methods):

$$T = 22\ ^\circ\text{C} : A_2 = 3.5(\pm 2) \times 10^{-4}\ \text{cm}^3\ \text{mol g}^{-2}$$

$$T = 53\ ^\circ\text{C} : A_2 = 5(\pm 2) \times 10^{-4}\ \text{cm}^3\ \text{mol g}^{-2}$$

The low positive values of A_2 indicate that the interactions between proteins, which essentially result from three contributions, hard-sphere, coulombic and attractive van der Waals interactions, are weakly repulsive under our conditions. By comparison, the second virial coefficient arising from hard-sphere interactions alone can be estimated theoretically. Des Cloiseaux & Janninck[47] have established the following relation between A_2 and R_g for uncharged macromolecules in solution:

$$A_2 = 4\pi^{3/2} \Psi N_A \frac{R_g^3}{M^2}$$

where N_A is the Avogadro number and M the macromolecule molar mass. Ψ is a dimensionless constant that depends on the molecular shape. For

a spherical shape, $\Psi \cong 1.619$.[48] With $R_g = 14$ Å, a value of 4.3×10^{-4} cm^3 mol g^{-2} is obtained for A_2, which is of the same order of magnitude as the experimental determinations. This suggests that under the ionic conditions used for the native NCS, the sum of all interactions is equivalent to hard-sphere interactions, which are weak at the low protein concentration used. The observed slight increase of the virial coefficient A_2 with temperature has already been reported in the case of lysozyme[49] and attributed to a decrease of the attractive dispersive forces with temperature, whilst repulsive interactions remain roughly constant. In any case, the NCS solution can be considered as ideal in the temperature range where the protein retains its native conformation. We may therefore reasonably suppose that the scattering pattern modifications observed during unfolding are due to protein structure changes and are not significantly altered by interaction effects.

The scattering pattern of the native protein at room temperature could then be analysed neglecting the interaction effects. The experimental curve $I(Q,c = 5$ mg ml^{-1}, $T = 22.8\,^{\circ}$C) was fitted by the curve representative of an ideal solution of native NCS calculated from the protein crystallographic co-ordinates (PDB file 1noa) using CRYSOL.[50,51] This program uses only two free parameters, the average displaced solvent volume per atomic group and the contrast of the hydration shell. The fit yields an excellent agreement over the whole angular range for a relative density of the hydration layer of 1.03, with an R_g value derived from the calculated curve of 14.0 Å, identical with the experimental determination (Figure 2(a)). All experimental data of essentially native protein (i.e. recorded at temperature up to 61.6 °C) display a similar agreement, even in the large-angle region. Whilst confirming the native character of the protein up to 61.6 °C, this perfect agreement testifies to the accuracy of our background subtraction. Since this particular step is the direct result of the experimental measurements, and not of a fitting procedure, we can reasonably assume that the background subtraction is performed accurately at higher temperature as well.

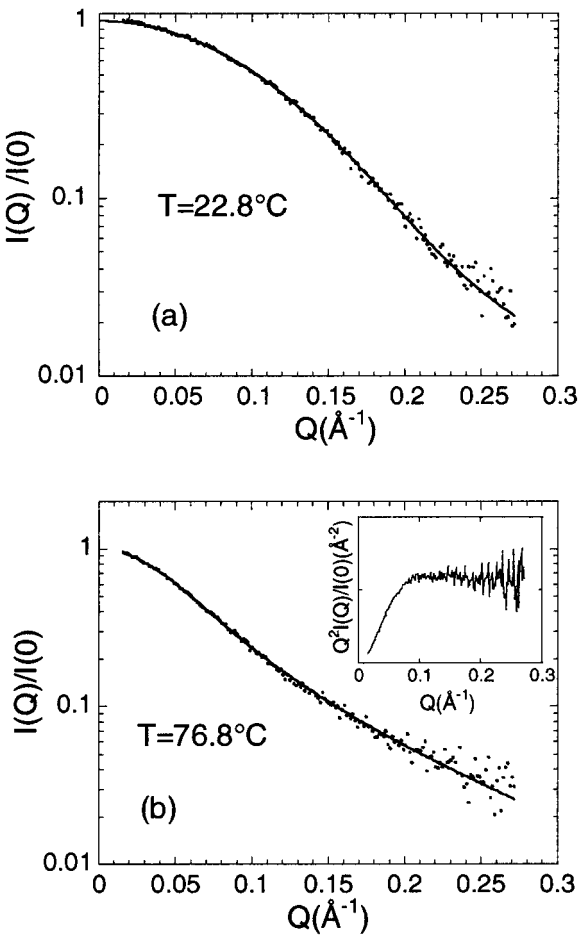

Figure 2. SAXS patterns $I(Q,c = 5$ mg ml$^{-1})$ of the two extreme conformations of NCS. (a) scattering pattern of native NCS at 22.8 °C. Dots, experimental data; continuous line, calculated pattern from the atomic co-ordinates. (b) Scattering pattern of unfolded NCS at 76.8 °C. Dots, experimental data; continuous line, fit using the Kratky-Porod model of a thick polymeric chain without excluded volume interactions. Inset: The Kratky plot ($I(Q)Q^2$ as a function of Q) exhibits a plateau at large Q values.

Completely unfolded protein

The experimental pattern measured at 76.8 °C presents at large Q the slow Q-dependence typical of an unfolded protein: its Kratky representation clearly displays a plateau at Q-values above 0.1 Å^{-1} (see the inset in Figure 2(b)). A scattering curve recorded at 80 °C appears to be essentially identical with that at 76.8 °C (data not shown). Since the former curve displays poorer statistics, only the 76.8 °C curve is used to analyse the unfolded state. The approximation to be used to derive the radius of gyration from the small-angle region of scattering curves differs according to whether the protein is native or unfolded (see Materials and Methods). The radius of gyration of the NCS unfolded state was therefore estimated by fitting the scattering pattern measured at 76.8 °C using the Debye formula applied in the range 0.02 Å$^{-1} < Q < 0.055$ Å^{-1}, which gave a value of 26.3(±0.5) Å. As in the case of the native protein, our first concern was to evaluate the effect of interactions. For practical reasons the second virial coefficient A_2 was not measured in our SAXS experiments, but it had already been evaluated in SANS experiments, under the same buffer conditions in ^2H_2O and was found to be very close to zero.[52] The virial of native proteins is known to be only slightly different in ^2H_2O and H_2O.[53] We assume here that the same applies for unfolded proteins. The negligible value of A_2 found by SANS means that the NCS scattering curve at 76.8 °C is not altered by interactions and that the

temperature-unfolded NCS can be considered as a polymeric chain in a theta solvent, i.e. with no excluded volume interactions.[54] The analytical model of a Kratky-Porod chain, with a finite value of the section radius (see Materials and Methods), can then be used to fit the data.

The excellent agreement between the experimental curve ($0.015\ \text{Å}^{-1} < Q < 0.2\ \text{Å}^{-1}$) and its fit by the Kratky-Porod model (equations (15) and (18)), even at the smallest Q-value, rules out any aggregation process (Figure 2(b)). The following values of the fitting parameters are obtained: $L = 320(\pm 70)$ Å, $b = 13.1(\pm 3)$ Å, $R_{g,c} = 4.15(\pm 0.3)$ Å. The radius of gyration value derived from equation (20) is 26.1 Å, nearly identical with that found using the approximation of the Debye's formula.

Transition zone

Radius of gyration. To evaluate accurately the average radius of gyration of the protein as a function of temperature during the unfolding transition, we have to determine which type of analysis (Guinier or Debye) is most appropriate. Indeed, the simultaneous presence of different states of the protein in the solution (compact, unfolded and possibly partially unfolded) makes this question more complicated than in the two simple and extreme cases of solutions with pure compact or pure unfolded states. A third derivation of the value of the radius of gyration is by means of the distance distribution function (see Materials and Methods). In a test case presented in the Appendix, all three methods were used to calculate the radius of gyration of synthetic scattering patterns $I_{sy}(Q)$ which are linear combinations (coefficients: $1 - f_U$, f_U) of the native and the unfolded protein scattering curves, simulating a two-state transition between a globular and an unstructured states. The comparison of the R_g values obtained by each of the three approaches with the true value provided us with a direct assessment of the way each method performed. Following the conclusion of this study, the average radius of gyration of NCS was determined as a function of temperature using the best approximation. The scattering curves measured at temperatures lower than 63.4 °C and fitted using the Guinier approximation are shown in Figure 3(a). As expected in this temperature range, $\ln(I(Q))$ is perfectly linear as a function of Q^2 at least out to $Q = 0.1\ \text{Å}^{-1}$, which corresponds to $QR_g = 1.4$. At higher temperatures, the graph of $\ln(I(Q))$ with Q^2 displays an increasing deviation from linearity (the patterns at $T = 65.1$ and 66.8 °C are also shown in Figure 3(a)). The fits using Debye's formula over the Q-range $0.02\ \text{Å}^{-1} < Q < 0.055\ \text{Å}^{-1}$ are shown in Figure 3(b) for temperatures above 65.1 °C. The absence of aggregation clearly appears from the perfect agreement of the fitting curves to the experimental data down to the smallest Q-values. No correction for interaction effects was performed on these curves, bearing in mind that these effects are quite small under the conditions used and even nil for the unfolded protein. Finally, the radius of gyration was derived from the corresponding distance distribution function $P(r)$ at each temperature.

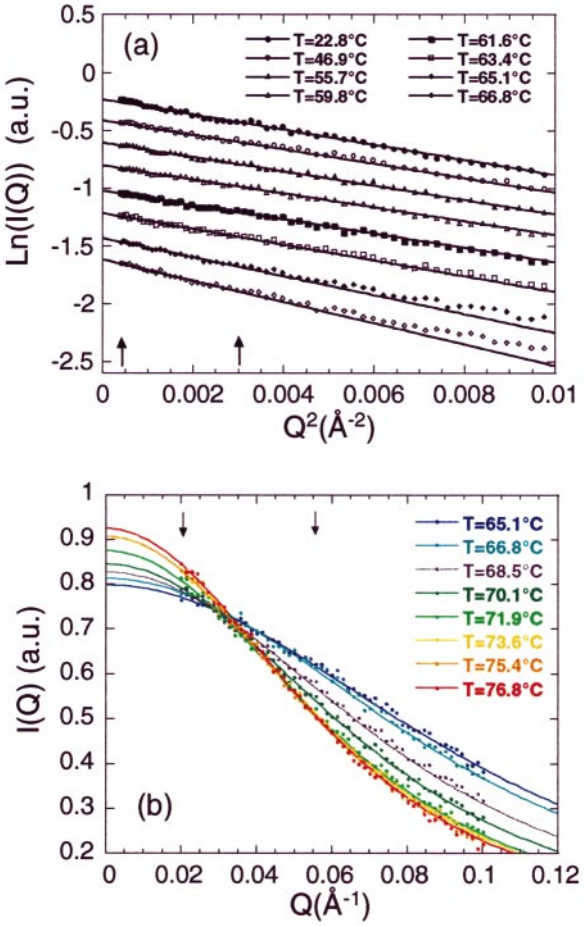

Figure 3. Determination of the apparent radius of gyration. (a) Guinier plots ($\ln(I(Q))$ as a function of Q^2) from data recorded between 22.8 °C and 66.8 °C. Below 63.4 °C, $\ln(I(Q))$ is perfectly linear as a function of Q^2 at least out to $Q = 0.1\ \text{Å}^{-1}$, which corresponds to $QR_g = 1.4$. At higher temperatures, an increasing deviation from linearity is observed from $Q = 0.06\ \text{Å}^{-1}$. (b) Fits, using Debye's formula, of data recorded between 65.1 °C and 76.8 °C. The arrows indicate the Q range used for the fits, $0.02\ \text{Å}^{-1} < Q < 0.055\ \text{Å}^{-1}$.

Figure 4(a) illustrates the variation of the square of the apparent radius of gyration $R_{g,app}$ as a function of temperature using the three methods. The radius of gyration of the protein remains practically constant up to 65.1 °C. The weak decrease from 22 °C to 60 °C is due to the slightly higher effects of concentration (5 mg ml^{-1}) at high temperature, as observed by comparing the two concentration series performed at 22 °C and 53 °C already mentioned. Beyond 65.1 °C, the average radius of gyration displays a sharp increase. The transition from the globular shape to the unfolded chain occurs within ca. 10 deg. C. Above 76.8 °C,

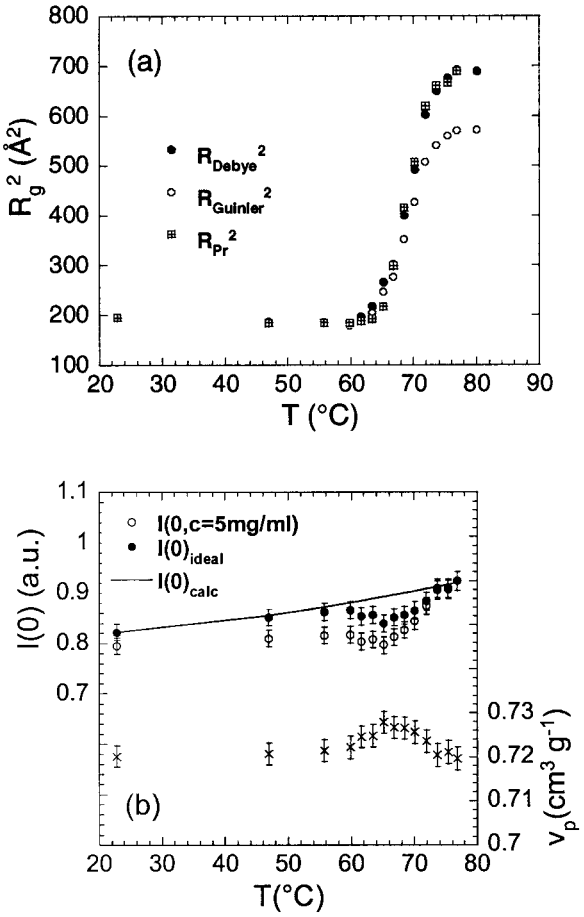

Figure 4. Temperature-dependence of R_g and $I(0)$. (a) Square of the radius of gyration. Open circles, values from Guinier's law. Filled circles, values from Debye's formula. Crosses in square, values from $P(r)$. (b) Intensity at the origin. Open circles, $I(0, c = 5$ mg ml$^{-1})$. Filled circles, after correction for interaction effect, $I(0)_{ideal}$. Continuous line, $I(0)_{calc}$ calculated assuming a constant partial specific volume of the protein. Crosses (right-hand scale), temperature-dependence of the partial specific volume of the protein as derived from the deviations between $I(0)_{calc}$ and $I(0)_{ideal}$.

$R_{g,app}$ remains constant. Confirming the conclusions given in the Appendix, the estimate of the radius of gyration yielded by the Guinier approximation is clearly smaller than that obtained from the Debye's formula as soon as the protein starts unfolding. The $P(r)$ derived values at 63.4 °C and 65.1 °C are significantly lower even than the Guinier estimate, in agreement with the observations made in the test case. With the exception of these two values where the Debye approximation should be retained (see Appendix), the differences between R_{Debye} and R_{Pr} appear to be much smaller than that observed in the test case, and well within the combined error bars. Both determinations are therefore likely to be close to the actual value.

Intensity at Q=0. Together with the radius of gyration, the analysis of the scattering curves in the small-angle range yields the value of the extrapolated intensity at zero angle, $I(0,c)$. Although the present SAXS experiment yields only a relative value of $I(0,c)$, not an absolute value, two pieces of information can be derived from the variation of $I(0,c)$ with temperature, namely the evolution of the average mass of the scattering objects and the evolution of the partial specific volume of the protein during unfolding.

Figure 4(b) shows the evolution of $I(0,c = 5$ mg ml$^{-1})$ with temperature, as extracted from the fits to the experimental data (Guinier approximation for $T < 62$ °C and Debye approximation for $T > 62$ °C). The absolute value of the second virial coefficient A_2 of NCS solutions is low in the native state and close to zero in the unfolded state. We assume that, during the transition, the apparent virial coefficient of the protein solution is a linear combination of those of the native and the unfolded forms of the protein. This very crude approximation is deemed to be acceptable in view of the very low values of A_2. This enables us to derive from the experimental $I(0,c)$ the intensity at the origin $I(0)_{ideal}$ corrected for the effect of interparticle interactions (equation 11)). The variation of $I(0)_{ideal}$ with temperature is also shown in Figure 4(b).

Besides, the mere effect of temperature on the solvent density leads to a variation of the intensity at zero angle easily computed using equation (9b). Assuming a constant value of 0.72 cm^{3}g^{-1} for $\bar{v}_p$ derived following the approach of Kharakoz,[52] the value $I(0)_{calc}$ can be evaluated from the tabulated variation of ρ_s with temperature. The decrease of the protein concentration associated with the decrease of the solvent density with increasing temperature has also been taken into account in this evaluation (c factor in equation (9b)). The variation of $I(0)_{calc}$ with T is plotted in Figure 4(b) on top of the $I(0)_{ideal}$ curve. Weak but systematic deviations are observed in the region of the transition between the calculated intensity $I(0)_{calc}$ and the experimental points. This can be attributed to various factors, which are discussed in the Discussion.

Distance distribution function. The $P(r)$ curves are plotted in Figure 5 for all temperatures. The shape of the $P(r)$ function remains essentially the same until 63.4 °C, beyond which the maximum extension D_{max} of $P(r)$ starts increasing. Beyond this point, the shape of $P(r)$ clearly changes, progressively leading to the profile of an unfolded chain at 76.8 °C. The shape of the $P(r)$ at $T = 65.1$ °C with its sharply ending tail appears to be somewhat unrealistic. This is because the proportion of unfolded or partially unfolded species is very low at this temperature, which makes it difficult to determine the $P(r)$ curve unambiguously (see Appendix).

The evolution of D_{max} with temperature is presented in the inset to Figure 5. Although the value

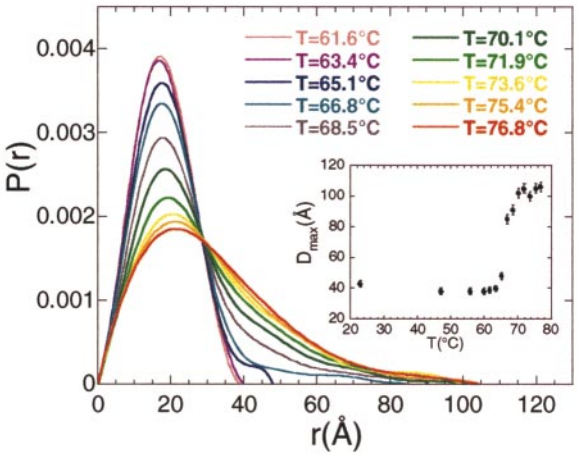

Figure 5. Distance distribution function of NCS at different temperatures across the transition scaled to the intensity at the origin. Note the unlikely shape of the tail at 65.1 °C reflecting the difficulty in accounting for the large distance side of the distribution in the case of a small fraction of extended conformation. Inset: Temperature-dependence of the maximum diameter, D_{max}.

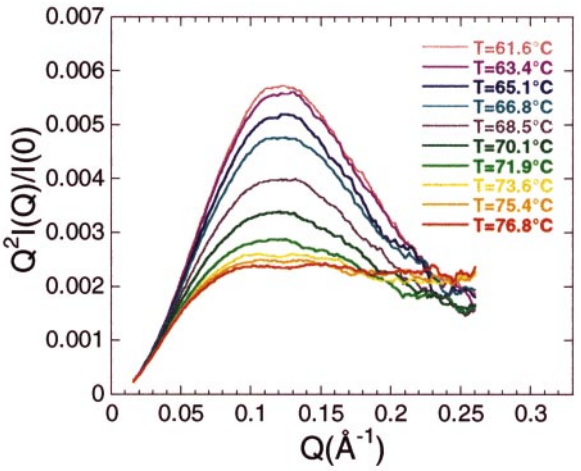

Figure 6. Kratky plots ($Q^2I(Q)$) as a function of Q) at different temperatures across the transition. Intensities have been linearly smoothed using a moving window for clarity. The deviations between the raw data and the smoothed values have been checked to be distributed randomly around zero.

of D_{max} is not determined with the same precision as R_g, its variation with temperature appears to be faster than that of R_g^2. D_{max} keeps close to 40 Å up to 65 °C, at which temperature it undergoes a sharp increase to reach its maximum value of about 105 Å above 70 °C.

Kratky representation. All normalised experimental curves, $Q^2I(Q,c)/I(0,c)$, for temperatures between 61.6 and 76.8 °C, are plotted in Figure 6. The curve at 61.6 °C displays the typical bell-shape of a native, globular protein. Beyond 63.4 °C, the peak amplitude progressively decreases, finally leading at 76.8 °C to a curve displaying a plateau for $Q > 0.1$ Å^{-1}, which characterises the completely unfolded protein. Note that the scattering curves do not cross in a single point during the transition. As stated in Materials and Methods, the absence of an isoscattering point unambiguously establishes that the solution contains intermediate states between the native and the completely unfolded. Besides, this absence of a common crossing-point is also visible, although less clearly, within the set of $P(r)$ curves in Figure 5.

To prove that the absence of an isoscattering point is not due to uncorrected interaction effects, the curves scaled to the intensity at the origin extrapolated at zero concentration were also plotted in the Kratky representation. The correction appears to be actually very small, and no isoscattering point is observed. In any case, an isoscattering point could be observed only at large angles, where interaction effects are negligible. All further analyses will thus be performed on the non-corrected curves, since the correction is small and depends on several approximations.

Analysis by singular value decomposition. An SVD analysis was performed on the experimental scattering curves using the Kratky expression scaled to the intensity at zero angle $Q^2I(Q,T_k)/I(0,T_k)$. Only the ten curves recorded at temperatures between 61.6 °C and 76.8 °C were considered. The reason for using a restricted range of temperature around the transition is developed in Materials and Methods: the SVD analysis is valid only in as much as the form factor of a given state does not vary over the temperature range explored.

The minimum number of species with linearly independent scattering curves necessary to reconstruct the whole data set was determined after calculating the χ^2 value given by equation (30) and by examining the shape of the basis vectors $\mathbf{u}_j(Q)$ together with the temperature-dependence of their respective coefficients $w_jb_j^k$. Figure 7 shows the first four basis vectors with the values of χ^2 as a function of the number p of basis vectors in the inset while the coefficients w_jb_j, $j = 1,2,3,4$, of the first four basis vectors are shown in Figure 8 as a function of temperature. It appears that at least three independent species are necessary to describe the transition, since the first three basis vectors and their associated coefficients display a meaningful Q or T-dependence, whilst the value of χ^2 drops to 1 for $p = 3$. In contrast, no significant correlation is observed between w_4b_4 and temperature, while the shape of $\mathbf{u}_4(Q)$ does not display any meaningful feature. We finally checked that the reduced residuals between the experimental curves and their reconstruction using the first three basis vectors are distributed randomly

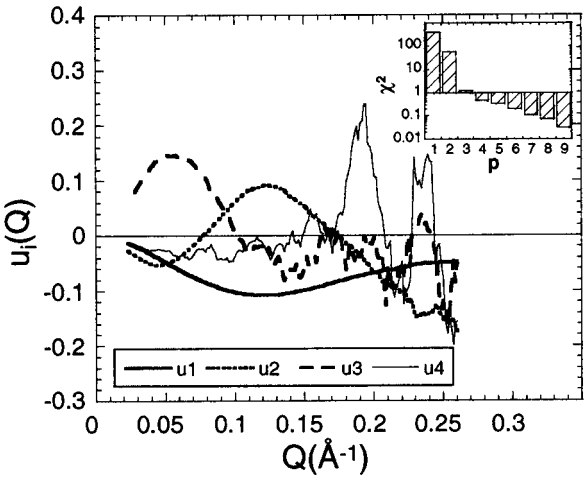

Figure 7. SVD analysis of SAXS patterns: first four basis vectors $\mathbf{u}_i$, $i = 1,2,3,4$ (unitary vectors). Inset: evolution of the χ^2 value (logarithmic scale) as a function of the number p of eigenvectors.

between +1 and −1. Therefore, three and only three species are detected by the SVD analysis ($L = 3$), meaning that, within experimental errors, an intermediate state between the native and the completely unfolded states appears during the transition. There might exist more than one intermediate state but, in this case, more precise data would be necessary to detect them.

Once the number of independent species is determined, the first three projection coefficients, $w_i b_i^k$, are fitted by equation (34) to yield the fractions f_N, f_I, and f_U of the native, intermediate and unfolded states, as a function of temperature together with the projection coefficients, p_j^I, of the intermediate state. From these coefficients, the scattering patterns $I_I(Q)$ of the intermediate state can be reconstructed using equation (33). Given the high number of fitting parameters, two constraints were imposed during fitting: the scattering patterns of the native and the unfolded state are imposed to be equal to the projections on the first three basis vectors of the curves recorded at $T = 61.6\,^\circ$C and $T = 76.8\,^\circ$C, respectively. The first constraint is justified, since the experimental curve measured at 61.6 °C is reproduced perfectly from the crystallographic coordinates of the native NCS. The second constraint is based on the identity between the scattering curves recorded at 76.8 °C and 80 °C, suggesting that the protein is already completely unfolded at 76.8 °C. However, as explained in Materials and Methods, these two constraints are not sufficient to define a single solution to this problem. We are left with a whole class that can be derived from any one of them by simple linear relationships (equation (37)).

A representative selection of this ensemble of curves, each of which is a possible scattering pattern of the intermediate state, is shown in Figure 9(a) using the Kratky representation. Although the curves exhibit a significant spread, all suggest that the intermediate state is largely unstructured. This is most easily seen by the values of the radius of gyration derived from the Debye approximation, which are all larger than 21.3 Å, to be compared with 14 Å and 26 Å, the respective values for the native and unfolded states. However, the protein in the intermediate state clearly presents elements of residual structure, as shown by the presence of a large peak at $Q = 0.11$ Å^{-1}. Note that the curves derived from equation (37) and presenting an apparent radius of gyration higher than 26 Å were considered to be physically meaningless and are not represented in Figure 9(a).

Figure 9(b) shows the evolution with temperature of the respective fractions of the native, intermediate and unfolded states. In all cases, the intermediate state is maximally populated between 68 °C and 72 °C, that is, in the neighbourhood of the half-transition temperature. Below 68 °C, the unfolded state is weakly populated, the solution mostly containing proteins in the native and intermediate states. Conversely, above 72 °C, the protein is predominantly in the intermediate and completely unfolded states.

Use of a thermodynamic model. The attempt to characterise the intermediate state with a minimal number of hypotheses leads us to the limited conclusion that this species must be unstructured. To go any further requires the introduction of a model restricting the range of scattering patterns compatible with the data. An approach has already been proven efficient in several cases,[42,44] which assumes the existence of a single thermodynamically well-defined state, denoted I, which is in equilibrium with the native (N) and unfolded (U) states with associated changes in free energy $\Delta G_{N \to i}$ ($i = $ I,U). Within this model, the coefficients f_i^k are entirely determined from the knowledge of the two Gibbs energies (see Materials and Methods), thereby considerably reducing the number of adjustable parameters. This calculation yields the fractions and scattering patterns shown in Figure 9(d) and (c), respectively. The scattering curve of the intermediate species, which is now unique, falls in the range previously obtained, while the fraction of intermediate species presents, as expected, a maximum around the transition temperature. Although worse than that observed before, the fit to the experimental data remains acceptable. However, the values derived for the free energies at ambient temperature, $\Delta G_{N \to I}$(300 K) and $\Delta G_{N \to U}$(300 K), are 16.5 and 43.8 kcal mol^{-1}, respectively, almost an order of magnitude larger than experimental determinations in similar systems.[56] This suggests that the hypothesis used to build the model, namely the existence of a thermodynamically well-defined intermediate state, is simply not valid in the case of the thermal unfolding of NCS.

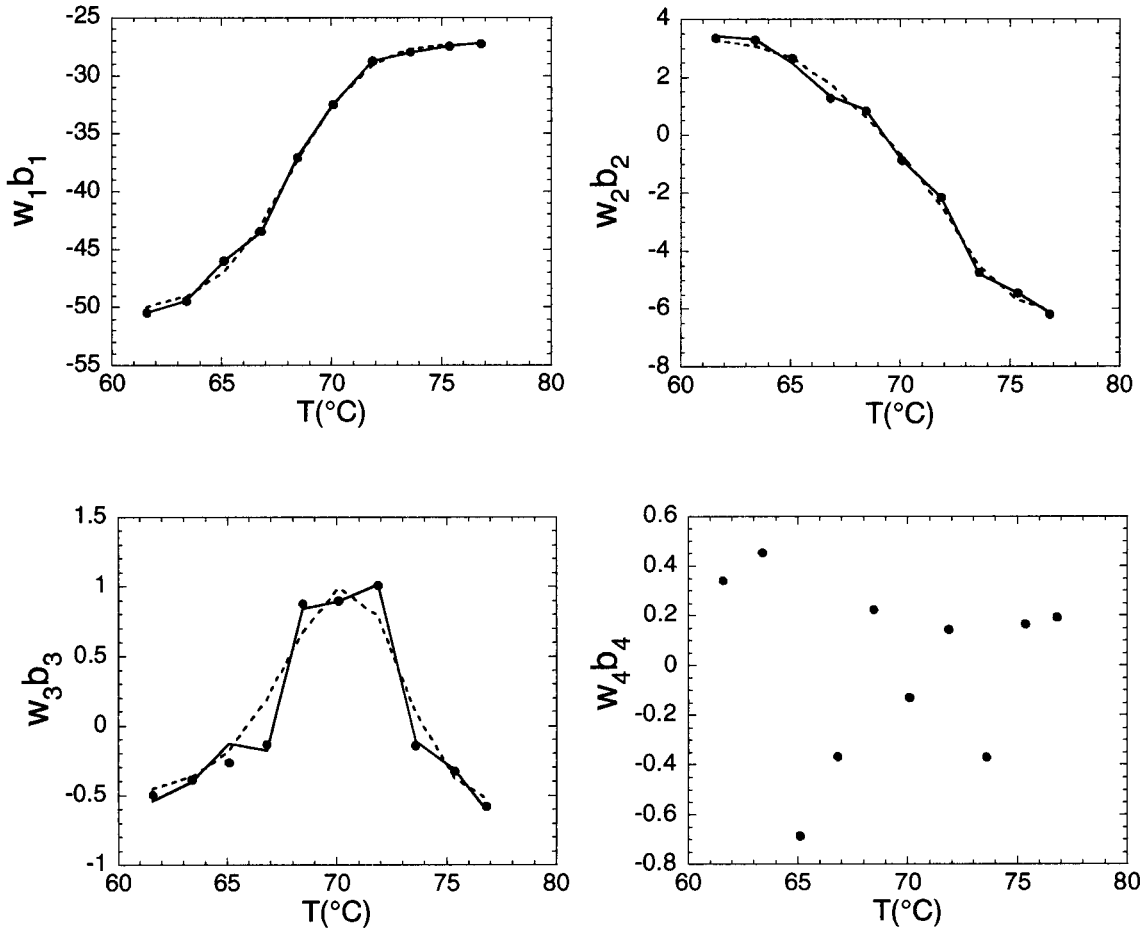

Figure 8. Temperature-dependence of the coefficients of the first four basis vectors. Dots, projections of the experimental Kratky profiles on the first four basis vectors. The projection on the fourth eigenvector varies randomly with temperature. The continuous line joins the points obtained using equation (35), which correspond to the three-state analysis with no thermodynamic model (see the text for details). Broken line, the same but using the thermodynamic model.

Discussion

Methodological comments

The monitoring of protein denaturation by small-angle X-ray scattering yields invaluable information about the global conformation of the various thermodynamic states involved. Moreover, the existence of intermediate states can be ascertained directly, with no need for a theoretical model used for data fitting. However, for the information to be useful, scattering intensities must be recorded over a large angular range and thoroughly analysed. Great care has to be taken in data reduction and analysis to obtain meaningful information on the structural characteristics of non-native states as well as on the existence of intermediate states. To start with, retrieving information on the conformation of a denatured protein from its scattering pattern requires a very accurate background subtraction. Regarding the native conformation of the protein, this is easily checked by comparing the experimental scattering pattern to the scattering curve computed from the crystal structure, when available, using the program CRYSOL.[50] Second, the influence of interaction effects on scattering curves must be evaluated and kept to a minimum. Some authors[57] extrapolate the scattering curves to infinite dilution in order to free the data from such effects. Although a valid option, this extrapolation is not always straightforward, and we therefore prefer, by an appropriate choice of the ionic conditions, to minimise repulsive coulombic interactions between proteins so that the second virial coefficient A_2 remains weak and the shape of the scattering pattern (recorded at a protein concentration lower or equal to 5 mg ml^{-1}) is not significantly affected by interaction effects. Ideally, the second virial coefficient should be slightly positive, that is the net resulting interaction should be weakly repulsive, so as to prevent proteins from aggregating during the denaturation process. The third comment regards the calculation of the apparent radius of gyration from the scatter-

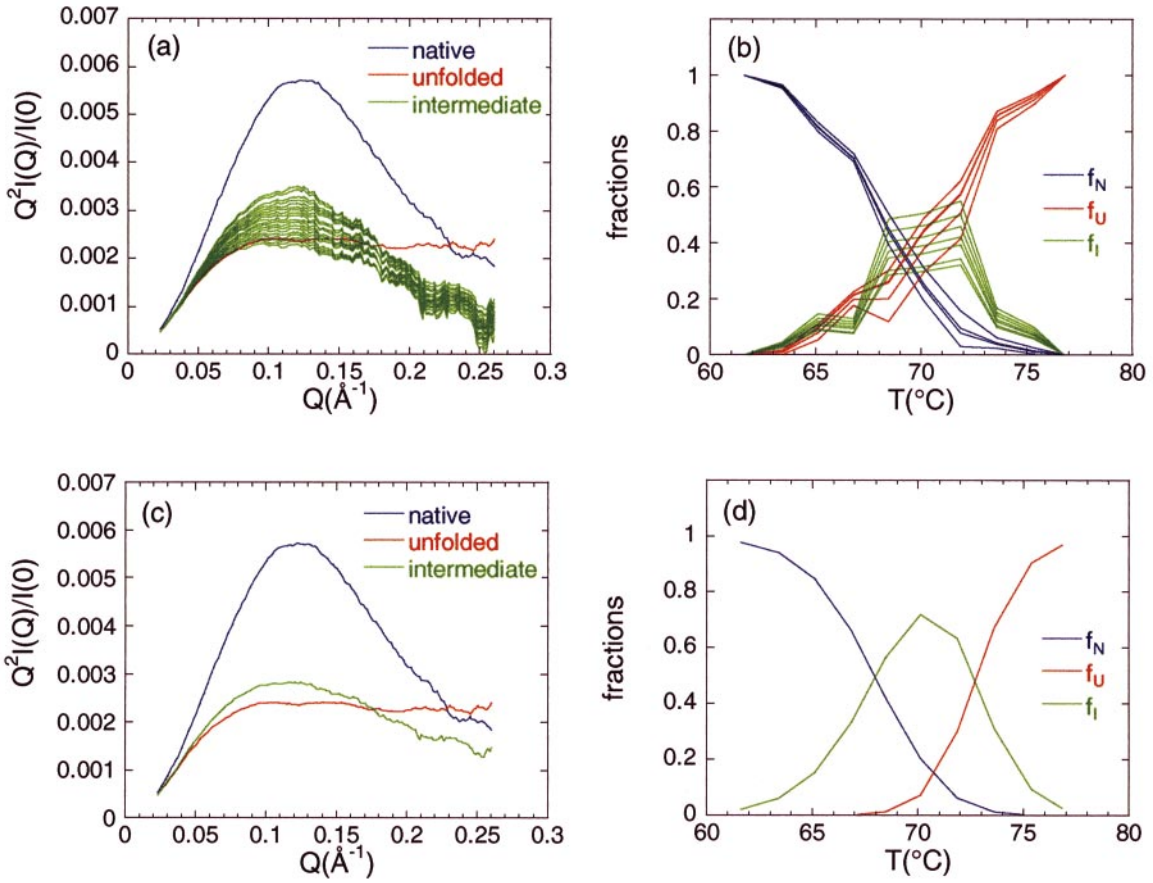

Figure 9. Scattering pattern and fractional concentration of the intermediate species with or without a thermodynamic model. No thermodynamic model: (a) class of scattering patterns of the intermediate state compatible with the experimental data; (b) fractional concentration of the native, fully unfolded and intermediate state. Seven sets of values are presented, spanning the whole range for the intermediate species. Note that several f_N and f_U curves are practically identical. Use of a thermodynamic model: (c) scattering pattern of the intermediate state; (d) fractional concentration of the native, fully unfolded and intermediate state.

ing patterns. In theory, the best approach remains the calculation of the $P(r)$ function, although its practical use is often not straightforward and shows its limits at the onset of the transition, when extended conformations are present in weak amounts. It is best supplemented by Debye's formula, which, in spite of the theoretical shortcomings mentioned in the Appendix, yields more accurate values than Guinier's law, provided the fit against experimental data is performed over a restricted angular range ($Q_{max}R_g < 1.4$). Indeed, the most generally used Guinier's law appears to give too small values of the radius of gyration from the beginning of the unfolding transition. Another consequence of the difficulties encountered in calculating the $P(r)$ function at the beginning of the transition regards the evolution of D_{max} with temperature. Unlike the square of the radius of gyration, the maximum extension of the protein is not an adequate parameter to follow the transition, since it does not vary linearly as a function of the fractions of the different species present in solution. In principle, the value of D_{max} is that of the most extended species detectable in solution. In the case of a purely two-state unfolding transition, D_{max} is actually expected to display a stepwise variation with temperature, changing steeply from its native state to its unfolded state value. However, it is shown in the Appendix that this is not the case in practice. The observation in Figure 5 of some intermediate values of D_{max} between 40 Å and 105 Å (one point at 48 Å and two points at 85 Å and 91 Å) can again be attributed to the difficulties mentioned in the test case to determine the $P(r)$ function at the onset of the transition. Therefore they cannot be considered as a proof of the existence of an intermediate conformation.

Finally, the use of isobestic points and of the SVD analysis calls for some words of caution. The scattering patterns must first be normalized by the intensity at the origin $I(0)$, so as to take into account the variation of the electron density contrast of the protein as a function of temperature (respectively denaturing agent concentration). The analysis must then be restricted to a narrow temperature (respectively denaturing agent concen-

tration) range in the immediate vicinity of the unfolding transition, so that the conformation of the completely denatured protein (as characterised by the parameters b and $R_{g,c}$) does not vary in any significant way. The SVD analysis is also best performed over a restricted temperature range, so that most patterns contain contributions from any putative intermediate species, thereby maximising the detectability of putative intermediate states. If too many patterns of either the native protein or the completely unfolded one are included, they will contribute excessively to the data set, with a risk of masking these very intermediate states one is looking for, unless an appropriate weighting scheme is used.

Description of the denatured state at 76.8°C

SAXS is a technique of choice to characterise the global conformation of an unfolded protein. We have shown that the scattering pattern of the denatured protein recorded at 76.8 °C is similar to that of the Kratky-Porod model of a stiff chain with a measurable radius of cross-section and no excluded volume interactions. We now discuss the values of the three basic parameters used in the model description: $R_{g,c}$, the apparent radius of gyration of the cross-section; L, the contour length of the chain; and the statistical length b, which is twice the persistence length of the chain.

The value of $R_{g,c}$ gives an estimate of the average transverse dimensions of the chain, but it is only a crude one, since it depends strongly on the structure and the chemical composition of the residues, as well as on the solvent structure around the chain. Interestingly though, the value derived from the fit of the model against the experimental data is 4.1(±0.3) Å, close to the average dimension of amino acid side-chains.

The contour length, L, of the chain represents its full extension length. Its value can be related to the level of unfolding of the polypeptide chain. Indeed, if the chain is completely unfolded, its contour length is given by the expression nl_0f, where n is the number of residues, l_0 is the distance between two adjacent C^α ($l_0 = 3.78$ Å) and f is a geometrical factor depending on the bond angles along the chain, which accounts for the fact that the completely extended polypeptide chain is not a straight line but displays zigzags. For a polypeptide chain, f is found equal to 0.95.[58] If the chain is not totally unfolded and displays residual structures, the contour length must be shorter than nl_0f. The experimental value is $L = 320(\pm 70)$ Å, to be compared to the value of $nl_0f = 100 \times 3.78 \times 0.95 = 360$ Å, in which the number of residues is taken to be $n = 100$, thereby excluding the two disulphide loops 37-47 and 88-93, which are preserved even above the temperature of transition. The comparison suggests that the level of unfolding of the protein in the denatured state at 76.8 °C is high, since the contour length is only 10 % shorter than the length of the totally unfolded chain. Besides, this is in agreement with the small value of the radius of gyration of the cross-section that is close to the average size of amino acid side-chains: indeed, if numerous clusters were present, the section of the protein would be larger than a single side-chain. However, no quantitative estimate of the content of residual structures can be derived from our SAXS patterns. Only NMR measurements can yield such an estimate (see, for example, some recent papers).[59–62] Furthermore, one should notice the large error on the value of L; this is due to the fact that the function describing the intensity scattered by a Kratky-Porod chain (equation (15)) depends mainly on the product Lb, with a specific dependence on the value of L restricted to the correction term with respect to Debye's formula (i.e. the second term in equation (15), the first one expressing Debye's formula). It is finally worth noting that the residual structures potentially present in the denatured protein at 76.8 °C are still present at 80 °C, since the two scattering patterns are superimposable within experimental uncertainties.

The statistical length b reflects the local rigidity of the chain. It corresponds to the description of the polypeptide chain as a polymer of rigid segments of length b, free to adopt random mutual orientations. The value of b is determined by short-range interactions, and therefore depends on the local molecular structure of the chain. For a given protein, the value of b can be determined theoretically, using a potential that takes into account all intramolecular contributions to the energy (see, for instance, Miller & Goebel[63]) with the possible addition of certain protein-solvent interactions as proposed by Rowe & Lopez-Pineiro.[64] Such conformational analyses allowed these authors to calculate the mean-square of the end-to-end distance, $\langle R^2 \rangle$ of several random coil proteins. The main result is that, for the investigated proteins, $\langle R^2 \rangle$ displays a roughly linear increase with the number of residues, with a slope between 70 and 90 Å^2/residue according to the model used. It has been shown that in the case of a Kratky-Porod chain, $\langle R^2 \rangle = Lb$,[65] hence $\langle R^2 \rangle = nl_0fb$ using the expression of L given above. The values given in the literature for the slope l_0fb yield a value of b between 19 and 25 Å according to the model used. The fit of the Kratky-Porod model against our experimental data gives a significantly lower value, $b = 13(\pm 3)$ Å. This difference could be due to the fact that the calculations have been performed keeping the temperature constant (room temperature), and cannot therefore take into account the effect of the temperature increase. It might also be due to the theoretical models themselves that make use of the values of the dipolar moments μ of the various amino acid residues, which are poorly known, while few experimental results are available to validate the theoretical calculations.[64]

It is finally of interest to compare the heat-unfolded and the chemically unfolded states of NCS. Indeed, the chemical unfolding of NCS using

guanidine-HCl as a denaturing agent has been studied by small-angle neutron scattering (SANS).[21] The value of the radius of gyration of the chemically unfolded state (above 5 M Gdn-HCl) is 33(±1) Å, significantly larger than the 26.3 Å of the thermally unfolded state. This difference expresses a genuine difference between the two conformation ensembles. While the thermally unfolded scattering pattern is that of a chain with persistent length and a measurable cross-section and no excluded volume interactions, the chemically unfolded state displays strong excluded volume interactions that cause the protein to adopt more expanded conformations. Similar differences have been reported for cytochrome *c*, a 104 residue long protein with no disulphide bridge. In this case, the value of the radius of gyration of the heat unfolded state is 23 Å, while values ranging from 28-32 Å are reported for the chemically unfolded state.[43,44]

$I(0)_{ideal}$ *versus T*

The experimentally observed temperature dependence of $I(0)_{ideal}$ has been compared with $I(0)_{calc}$, calculated from equation (9b) and assuming a constant value of the partial specific volume of the protein throughout the transition. Significant and systematic discrepancies are observed, which can result from several factors. First, slight changes in the sample concentration cannot be ruled out. However, these variations are small, of the order of 2-3 % at most, as estimated by measuring again the scattering pattern of the stock solution at room temperature at the end of the experiment. These uncertainties have been included in the error bars on $I(0)_{ideal}$ given in Figure 4(b) and are clearly too small to account for the discrepancies between $I(0)_{ideal}$ and $I(0)_{calc}$. Second, interparticle interactions in a mixture of proteins are complex and likely to change during the unfolding process. The linear temperature-dependence assumed for the virial coefficient A_2 through the transition might well be inadequate. Nevertheless, in view of the low absolute value of A_2, these effects are most likely to be weak, and should not significantly modify the value of the extrapolated intensity $I(0)_{ideal}$. In conclusion, we think that the variation with temperature of $I(0)_{ideal}$ is mainly due to the variation of the partial specific volume of the protein during the unfolding transition. Ascribing the discrepancies between experimental and calculated values of $I(0)_{ideal}$ to the sole effect of $\overline{v}_p$ changes, values of the apparent $\overline{v}_p$ can be derived from equation (9b) (see Figure 4(b)). The amplitude of the variation of $\overline{v}_p$ appears to be weak, of the order of 1 %, in agreement with the comment from Chalikian and colleagues that ''in general, protein denaturation is accompanied by small (close to zero) negative or positive volume changes''.[66,67]

Intermediate states

The absence of isoscattering point common to all scattering patterns recorded as a function of temperature strongly suggests that the thermal unfolding process of NCS involves at least one intermediate state. This is demonstrated by the SVD analysis, which shows that three components are necessary to reconstruct the scattering patterns, while the contribution from the fourth one lies within experimental uncertainties. This intermediate state could be characterised only partially, and appears to be very unstructured according to the high value of the radius of gyration, always larger than 21 Å, to be compared with a value of 14 Å for the native protein and 26.2 Å for the unfolded protein at 76.8 °C. However, all Kratky profiles of the ensemble of possible patterns of the intermediate state are significantly different from the profile of the unfolded protein typical of a Kratky-Porod chain. They display a clear maximum at a Q value close to 0.11 $Å^{-1}$, while the unfolded pattern displays a plateau for Q-values larger than 0.1 $Å^{-1}$. Finally, the values of D_{max} found for the intermediate species are larger than 90 Å, as compared to 39 Å for the native protein and 106 Å for the completely unfolded protein, also supporting the view of a very extended conformation.

No unique pattern could be determined for the detected intermediate conformation of NCS. In a similar study of the chemical denaturation of cytochrome *c*, Segel and colleagues succeeded in determining this curve using the thermodynamic approach that failed with NCS.[44] However, they further comment in the discussion of their results that, by increasing the accuracy of measurements and the number of scattering curves recorded in the transition range, more components would most likely be required to account for this improved data set, corresponding to the detection of as many new intermediate states. In other words, the intermediate state is just their best description, given the available data, of an ensemble of denatured states. This is a way of reconciling the results of the SVD analysis with the so-called ''new view'' of protein (un)folding. Indeed, whilst the former can detect and, with the addition of constraints from a model, characterise only one or several discrete species, the latter handles an ensemble of conformations corresponding to different regions of the energy landscape of the protein, some of which are separated by a pass, i.e. by a transition and an energy of activation.[68] In our case, the failure of the simple thermodynamic model suggests that there is more than the unique intermediate structure required by the SVD analysis of our data. Formally, there could be discrete intermediate species at equilibrium, or an ensemble of conformations differently populated according to the temperature. In the latter case, a third state would be a weighted average of all non-native and not completely unfolded states of the protein but, since the weights are changing with the temperature, no

meaningful curve is likely to be derived by a global analysis using a simple model like that used here. This could explain why our attempt met with a failure. Such a model of a temperature-dependent distribution of (non-native) states is reminiscent of that proposed by Griko and colleagues to describe the thermal denaturation of α-lactalbumin as studied by DSC.[69,70] This model of cooperative hierarchy considers thermal denaturation as a transition from the native state towards a particular distribution of denatured molecules with a variable amount of residual structure. The nature of the distribution depends on the solvent used: under some conditions, the major species present during the transition might be considerably unstructured, which makes the thermal unfolding transition look like a two-state transition. When the temperature increases, the center of the distribution gradually shifts towards the completely unfolded state. A similar mechanism has been used by Hirai and colleagues[71] to describe the thermal unfolding of lysozyme monitored by SAXS and DSC.

Following these general comments on the unfolding process, we wish to conclude with some considerations more specific to NCS. As mentioned in the introduction, NCS possesses an Ig-like fold. Hamill and colleagues analysed the structure of the transition state in the folding of the third fibronectin type III domain of human tenascin, TNfn3, and compared it to that of other Ig-like proteins.[10] They reached the conclusion that this state presents a folding core centred around amino acid residues I20, Y36, I59 and V70, which belong to the central strands B, C, E and F. In order to compare these results with NCS, Adjadj and colleagues (unpublished results) aligned the NCS sequence with that of fibronectin on the basis of structure homology. Residues V21, Q36, L67 and V95 of the NCS sequence appear to correspond to I20, Y36,I59 and V70 of TNfn3, respectively. It is remarkable that Adjadj and colleagues showed, based on the study of the methyl group motions of apo-NCS by ^{13}C NMR relaxation, that these four residues were part of a hydrophobic cluster. It can be assumed reasonably that this cluster is present in the transition state of NCS. Some preliminary calculations performed on models satisfying these constraints yield a value of the radius of gyration lower than 21 Å, the smallest value for the radius of gyration of the intermediate state compatible with our data. Therefore, this suggests that the ensemble of non-native states are less structured than this putative transition state.

Conclusion

NCS, a small β-protein with an Ig-fold, presents an unexpectedly complex unfolding transition with temperature. The process involves at least an intermediate state, as suggested by various observations and proven by the SVD analysis of the set of SAXS curves. Furthermore, it cannot be described satisfactorily by the classical thermodynamic model of three thermodynamically well-defined states. This favours a description in terms of an ensemble of intermediate states leading to the fully unfolded state, the population of which gradually changes with temperature. We have shown that the analysis of scattering patterns of a mixture of states is not straightforward, and that great care must be exerted to derive meaningful and reliable information. However, provided one is cautious in avoiding the pitfalls pointed out here, the potential contribution of SAXS to the study of the folding problem is unique and perfectly complements the wide range of spectroscopies commonly put to use.

Materials and Methods

Wild-type neocarzinostatin was expressed in *Escherichia coli* containing a synthetic gene coding for neocarzinostatin inserted into the expression vector pET12A (NOVAGENE®). The proteins were purified as described.[72]

Differential scanning calorimetry

The thermal stability and reversibility of heat denaturation were studied by scanning calorimetry on a DSC (Microcal Corp®). DSC measurements were made using a 3 mg ml^{-1} neocarzinostatin solution dialysed overnight against 60 mM phosphate buffer (pH 7.0). Buffer solution from the dialysis bath was used as a reference. All solutions were degassed just before loading into the calorimeter. Calibrations were done according to the manufacturer's procedures. Specifically, DSC calibration was performed by measuring the deflection of the baseline after a heat pulse of 5 mcal min^{-1} for five minutes on a water baseline scan. Temperature calibration was obtained by measuring the melting temperature of standards provided by the manufacturer and consisting of small amounts of pure paraffin hydrocarbons sealed in steel capillary tubes.

The heat capacity of the solvent alone was first subtracted from that of the protein sample. The analysis was performed on these corrected data using a cubic spline as a baseline below the transition peak. Thermodynamic parameters ΔH_{cal} and ΔH_{vH} were obtained by a least-squares fit of the following equation against the data:

$$\Delta C_p(T) = \frac{K_D(T)\Delta H_{cal}\Delta H_{vH}}{[1 + K_D(T)]^2 RT^2} \quad (1)$$

where K_D is the equilibrium constant for a two-state process, ΔH_{cal} is the measured enthalpy, corresponding to:

$$\Delta H = \int_{T_I}^{T_F} C_P(T)\mathrm{d}T \quad (2)$$

and ΔH_{vH} is the enthalpy calculated on the basis of a two-state process.

If the measured transition corresponds to a two-state process, the values of the two enthalpies ΔH_{cal} and ΔH_{vH}, are equal. Any difference between these two values implies either the presence of intermediates or a multimolecular process.

Fluorescence measurements

For fluorescence measurements, the sample was identical with that used for DSC measurement, except that protein concentration was 5 μM (0.5 mg ml^{-1}). The tryptophan fluorescence of the protein was measured on an Aminco SLM 8000 fluorimeter at various temperature (heating rate 1 K min^{-1}). The fraction of unfolded protein, f_u, at each temperature was calculated from the wavelength of maximum emission fluorescence λ using the standard equation:

$$f_u = \frac{\lambda_N - \lambda}{\lambda_N - \lambda_D} \tag{3}$$

where λ_N and λ_D represent, respectively, the wavelength of maximum emission fluorescence of the native and denatured species. ΔH and T_m (mid-point transition temperature) were calculated from the transition curves using the standard equation:

$$\frac{\partial \ln K}{\partial \frac{1}{T}} = -\frac{\Delta H}{R} \tag{4}$$

where K is defined as:

$$K = \frac{f_u}{1 - f_u} \tag{5}$$

and R is the gas constant. ΔH is assumed to be temperature-independent. T_m is defined as the temperature at which $\ln K = 0$.

Small-angle X-ray scattering measurements

NCS solution in 60 mM phosphate buffer (pH 7.0), was centrifuged for ten minutes at 10,000 rpm prior to X-ray analysis to eliminate all aggregates. NCS concentration was measured by UV absorption ($\varepsilon_{280} = 1.29$ mg ml^{-1} cm^{-1}). A stock solution was prepared at a final concentration of 5 mg ml^{-1}, stored at 4 °C and directly used for the experiments. At each temperature, the sample was incubated in the X-ray cell for about 15 minutes before the measurements. For two temperatures (22 °C and 53 °C), samples at different concentrations (2.5, 5, 7.5 and 10 mg ml^{-1}) were also analysed to investigate the concentration effects.

Scattering data were recorded on the small-angle X-ray scattering instrument D24 at LURE (Laboratoire pour l'Utilisation du Rayonnement Electromagnétique, Orsay, France) using the radiation emitted by a bending magnet of the storage ring DCI. The wavelength λ was selected by a bent Ge(111) monochromator and adjusted to 1.488 Å, calibrated by the nickel absorption edge. X-ray patterns were recorded using a linear position-sensitive detector filled with a 90% Xe/10% CO_2 gas mixture. The sample-to-detector distance was 1320 mm corresponding to the scattering vector range: $0.015\ Å^{-1} < Q < 0.27\ Å^{-1}$ (where $Q = 4\pi\sin\theta/\lambda$, 2θ is the scattering angle). The sample was placed in a quartz capillary temperature-controlled (better than ±0.1 deg. C) *via* circulating water. The air-scattering was virtually eliminated by inserting the cell in an evacuated beam path.[73] During the unfolding process, proteins display an increased sensitivity to X-ray irradiation (i.e. to free radicals). To avoid radiation-induced damage, the plunger of the syringe connected to the cell was coupled to an automatic drive. The translation of the sample through the X-ray beam was adjusted in such a way that protein molecules would be exposed to the X-rays for less than two minutes, while remaining at constant temperature. Several successive frames (usually eight) of 200 seconds each were recorded for both the sample and the corresponding buffer. Each frame was carefully inspected to check for any X-ray damage (none was found) before calculating the average intensity and the associated experimental error.

Each scattering spectrum was corrected for the detector response and scaled to the transmitted intensity, using the scattering intensity from a reference carbon-black sample integrated over a given angular range. At each temperature the scattering from the buffer was measured and subtracted from the corresponding protein sample pattern.

Analysis of the background-corrected scattering functions *I*(*Q*)

The theoretical formalism applied to analyse SAXS patterns depend on the type of object present in the sample. Thus it may differ according to whether the protein is native or unfolded. For instance, in the case of chemically denatured phosphoglycerate kinase,[74] the Guinier approximation usually applied to the scattering intensity $I(Q)$ has been proven unable to give a correct value of the radius of gyration, while the polymer formalism could provide a good description of the experimental patterns. This point is further developed now.

Non-interacting particles

Considering a solution of N non-interacting identical particles per unit volume, the scattering intensity per unit volume can be expressed as:

$$I(Q) = N\left\langle \left| \int_v (\rho(\vec{r}) - \rho_s) e^{i\vec{1}.\vec{r}} d^3r \right|^2 \right\rangle \tag{6}$$

where v is the volume of the particle, $\rho(\vec{r})$ its electronic density and ρ_s the electronic density of the solvent. The brackets represent spherical averaging. At small Q, this expression can be expanded in powers of Q and the well-known Guinier approximation[75] can be derived:

$$I(Q) = I(0)\exp\left(-\frac{Q^2 R_g^2}{3}\right) \tag{7}$$

where the radius of gyration R_g is given by:

$$R_g^2 = \frac{\int_v r^2(\rho(\vec{r}) - \rho_s) d^3r}{\int_v (\rho(\vec{r}) - \rho_s) d^3r} \tag{8}$$

and the intensity at $Q = 0$ may be expressed as:

$$I(0) = N\left((m_p - \rho_s V_p)\frac{M}{N_A}\right)^2 \tag{9a}$$

where N_A is the Avogadro number, M is the protein molar mass, m_p is the number of electrons of the dry protein and $\bar{v}_p$ is the protein partial specific volume, i.e. the volume occupied per unit mass by the protein within the solvent. The partial specific volume of the protein $\bar{v}_p$ takes into account the contribution of the hydration

water presenting an electron density contrast with bulk water.

Alternatively, $I(0)$ can be rewritten as:

$$I(0) = \frac{cM}{N_A}\left(\frac{m_p N_A}{M} - \rho_s \bar{v}_p\right)^2 \quad (9b)$$

where c is the protein concentration (w/v) of the solution.

The Guinier approximation (equation (7)) is valid only in as much as the contribution of the next term in the power expansion of $I(Q)$ is negligible. The coefficient of the Q^4 term involves higher-order moments of the distribution of the electron density contrast $\rho(r) - \rho_s$ associated with the protein. When proteins have a compact structure, the Q domain over which the law is valid is large enough for practical use. For instance, in the case of an ellipsoïd, this Q domain may extend to slightly more than $QR_g = 1$, giving an R_g value with about 1 % accuracy. In practice, ln ($I(Q)$) is fitted by a straight line as a function of Q^2 for $Q < 1/R_g$, yielding $I(0)$ and R_g in a straightforward way.

However, if proteins are unstructured enough to resemble a polymeric chain, distances between elements of the protein can be very large, so that the Q^4 term becomes significant for Q values small enough to make the Guinier approximation of no practical use.[65] It has been shown[74] in that case, that the scattering function $I(Q)$ is well described in the domain $Q < 3/R_g$ by the Debye equation:[76]

$$\frac{I(Q)}{I(0)} = \frac{2}{x^2}(x - 1 + e^{-x}) \quad (10)$$

where $x = Q^2R_g^2$. This relation is the correct analytical expression for an infinitely thin Gaussian chain, with no persistence length (i.e. with no rigidity due to short-range interactions between monomers) and with no excluded volume (i.e. with no long-range interactions between monomers). However, it remains valid over a restricted Q range in the case of an excluded volume chain or a chain with persistence length.

Interacting particles

In the case of interacting protein molecules, the scattering patterns are modified by the effects of the positional correlations due to interactions. The above expressions of scattering intensity must be modified. In particular the concentration dependence of the intensity at the origin is given by the following expression:

$$I(0, c) = \frac{I(0, 0)}{1 + 2A_2Mc} \quad (11)$$

where A_2 is the second virial coefficient and $I(0)_{ideal}$ is proportional to c according to equation (9b). The intensities in the small-angle region used for R_g determination are also modified and yield an apparent radius of gyration $R_g(c)$.

From a series of measurements carried out as a function of protein concentration, values of $I(0,c)$ are determined. Equation (11) is then used to extrapolate $I(0,c)$ to infinite dilution, yielding the A_2 coefficients together with the contrast factor k, where $I(0)_{ideal}$=cK.

Distance distribution function

The previous analysis concerns the small Q domain of the scattering patterns, and essentially yields information about the protein compactness, through the value of its radius of gyration. Additional information about the particle conformation can be extracted by calculating the distance distribution function $P(r)$, which is the Fourier transform of the entire scattering curve $I(Q)$:

$$P(r) = \frac{r^2}{2\pi^2}\int_0^\infty I(Q)\frac{\sin(Qr)}{Qr}Q^2 dQ \quad (12)$$

$P(r)$ is zero for distances r larger than the maximum diameter D_{max} of the particle. The $P(r)$ function was calculated from the measured intensity $I(Q)$ using the program GNOM.[77] The value of D_{max} is fixed by the user in a recursive way until $P(r)$ presents a smooth shape. In the cases where some ambiguity remains about the determination of D_{max}, we systematically choose the highest possible value above which the $P(r)$ function exhibits fluctuations. The radius of gyration and the intensity at the origin are derived from $P(r)$ using the following two expressions:

$$R_g^2 = \frac{\int_0^{D_{max}} r^2 P(r) dr}{2\int_0^{D_{max}} P(r) dr} \quad (13)$$

and:

$$I(0) = 4\pi \int_0^{D_{max}} P(r) dr \quad (14)$$

This alternative evaluation of R_g and $I(0)$ is of interest because it is independent of any model. The comparison of the resulting values with the previous determinations using Guinier or Debye laws therefore provides a useful cross-validation of the results.

Theoretical model for an unfolded protein

The Debye model introduced above (equation (10)) describes the scattering intensity from an ideal (or Gaussian) chain, which means an infinitely thin and long chain with no excluded volume interactions. This model may provide a satisfactory first approximation of an unfolded polypeptide chain and, in particular, may give its radius of gyration with good accuracy. However, more realistic models are required to gain information about the actual nature of the chain.

The first model that can be tested is that of a chain with persistence length (or Kratky-Porod chain),[78] which accounts for local rigidity. In a polypeptide chain, this rigidity results from the polypeptide bond itself and from the restricted angular range between two adjacent peptides, but further constraints may arise from the occurrence of possible elements with residual structure. Two parameters define this type of chain, its contour length L and its statistical length b, which is twice the persistence length. Sharp & Bloomfield[79] have given an approximate expression for the form factor of a Kratky-Porod chain with a finite length:

$$\frac{I(Q)}{I(0)} = P^{SB}(Q) = \frac{2}{x^2}(x - 1 + e^{-x}) + \frac{b}{L}\left[\frac{4}{15} + \frac{7}{15x} - \left(\frac{11}{15} + \frac{7}{15x}\right)\exp(-x)\right] \quad (15)$$

with $x = Q^2Lb/6$. The radius of gyration R_g is then given by:

$$(R_g^{SB})^2 = b^2\left[\frac{y}{6} - \frac{1}{4} + \frac{1}{4y} - \frac{1}{8y^2}(1 - \exp(-2y))\right] \quad (16)$$

where $y = L/b$.

The expression (15) is valid only for $L/b > 10$ and within the domain $Q < 3/b$. In the domain $3 < Qb < 7$, the correct representation of a Kratky-Porod chain has been derived by des Cloiseaux.[80] Finally, for higher values of Q ($Q > 7/b$) the asymptotic trend of the form factor is the well-known behaviour of a rod:

$$\frac{I(Q)}{I(0)} = \frac{\pi}{QL} + \frac{4}{3Q^2bL} \quad (17)$$

which displays a characteristic decrease as $1/Q$. However, in typical SAXS experiments, the latter Q range is hardly reached.

Besides the local rigidity of the main chain, the existence of lateral chains has to be accounted for. To this aim, one may use the model of a thick filament, in which the thickness of the polymer chain is characterised by the radius of gyration $R_{g,c}$ of the cross-section.[81] The form factor is then expressed as:

$$\frac{I(Q)}{I(0)} = P^{SB}(Q)\exp\left(-\frac{Q^2R_{g,c}^2}{2}\right) \quad (18)$$

Equation (18) is valid only within the Guinier domain of the section, i.e. $Q < 1/R_{g,c}$. The radius of gyration R_c of the cross-section is defined as:[81]

$$R_{g,c}^2 = \frac{\int \vec{x}^2(\rho(\vec{x}) - \rho_s)d^2x}{\int(\rho(\vec{x}) - \rho_s)d^2x} \quad (19)$$

where $\vec{x}$ is a vector perpendicular to the chain axis, while the integrals are performed over the section of the polymer. The radius of gyration of a thick filament is then:

$$R_g^2 = (R_g^{SB})^2 + \frac{3}{2}R_{g,c}^2 \quad (20)$$

The previous expressions constitute the most realistic and simple model that describes the scattered intensity by an unfolded polypeptide chain without excluded volume effects, i.e. with no long-range repulsion between residues. Within the polymer theory framework, this case corresponds to the ''theta solvent'' case.[54] If excluded volume interactions are present (i.e. the ''good solvent'' case), which is the case for any chemically denatured protein, using for instance guanidinium chloride, more sophisticated models have to be used to describe the scattered intensity,[52] like that proposed by Pedersen & Schurtenberger.[82]

Fitting the experimental curves $I(Q)$ with equations (15) and (18) allows the determination of the three parameters L, b and $R_{g,c}$. The first two characterise the chain structure. The third one might be less relevant to the chain structure, since it depends on the local contrast $\rho(\vec{x}) - \rho_s$, as shown in equation (19). Therefore, for a given chain, with given values of b and L, the radius of gyration of the cross-section $R_{g,c}$ may vary if ρ_S itself strongly varies, thus modifying the scattering function $I(Q)$, especially its behaviour at high Q.

Kratky plots

Kratky plots (i.e. $I(Q)Q^2$ as a function of Q) of scattering intensities are a representation of particular value in protein folding studies. Indeed, each type of model structure (spherical particle, ideal Gaussian chain, Porod-Kratky chain, etc.) can be identified easily by the characteristic profile of the associated Kratky plot. Thus, the intensity scattered by a globular protein behaves approximately as $1/Q^4$ in the high Q domain,[83] yielding a bell-shaped Kratky plot with a well-defined maximum. Conversely, the Kratky representation of an ideal Gaussian chain presents a plateau in the high Q domain, since $I(Q)$ varies as $1/Q^2$, according to equation (10). Finally, in the case of a Kratky-Porod chain with no thickness, the Kratky plot also displays a plateau, since $I(Q)$ varies approximately as $1/Q^2$ (equation (15)) over a specific Q range but, at higher Q, it is followed by a monotonic increase, characteristic of the scattering by a rod (equation (17)).

Based on these observations, assertions can be found in the literature that the intensity scattered by a completely unfolded protein must display the latter kind of behaviour and, in particular, that the absence of increasing intensity at large Q in a Kratky representation is a proof of incomplete unfolding. In fact, the relationship between the general shape of the Kratky plot and the nature of an unfolded chain is complex. Within the framework of the Kratky-Porod chain model, a variety of shapes can be observed: no intensity increase at high Q, absence of a plateau, or even an intensity decrease at high Q, all of them describing a completely unfolded chain. To be more explicit, we may first stress that the Q domain of the plateau depends on the values of b and L (equation (15)) , and might be severely limited in some cases. Moreover, if interactions of excluded volume are present, the plateau disappears, since the asymptotic behaviour of $I(Q)$ does not follow a $1/Q^2$ law anymore but $1/Q^{1/\nu}$ instead, with $\nu = 0.588$.[84] Second, the rod-type behaviour at high Q is predicted only for $Q \gg 7/b$, which, for $b = 13$ Å (the value found for NCS) corresponds to $Q \gg 0.5$ Å^{-1}, a value rarely reached in actual measurements. Third, the finite value of the radius of gyration of the cross-section may induce an extra decrease of $I(Q)$ (equation (18)), thereby masking the rod-type behaviour. In conclusion, the appearance of the Kratky plot of a Kratky-Porod chain of finite section depends on the relative values of b, L and R_c, and on the explored Q-range.

Intensity scattered by a solution containing several species

The previous section deals only with monodisperse solutions. During protein unfolding, at least two species, the native and the completely unfolded protein, are present simultaneously in solution. Assuming the interactions between particles to be negligible, i.e. the solution to be ideal, the total scattered intensity is the weighted sum of the intensities scattered by the various species:

$$I(Q,T) = \sum_i f_i(T) I_i(Q) \quad (21)$$

The f_i are the molar fractions of species i at a given temperature and the $I_i(Q)$ are the respective scattering intensities. In fact, the $I_i(Q)$ also vary during the unfolding process, since the contrast between proteins and solvent is modified by the temperature-dependence of water density. Hence, we may write:

$$I_i(Q,T) = c(T) K(T, \bar{v}_{p_i}) P_i(Q) \quad (22)$$

where $P_i(Q)$ is the form factor of species i ($P_i(0) = 1$) and $K(T,\bar{v}_{p_i})$ is the contrast factor, given by equation (9) and where the T dependence of the global protein concentration (w/v) is made explicit.

If we suppose that the differences between the partial specific volumes of each species are negligible, the contrast factor $K(T, \bar{v}_{p_i})$ depends only on temperature and equation (22) may be rewritten as:

$$\frac{I(Q,T)}{I(0,T)} = \sum_i f_i(T) P_i(Q) \quad (23)$$

This equation can be put to use to determine the scattering pattern of an unknown species together with the fractional concentrations of all species in solution using a least-squares fitting approach (see below). This requires the availability of experimental scattering patterns recorded as a function of temperature and their extrapolation to the origin $I(0,T)$. However, it is worth emphasizing that the derivation of equation (23) from the equation (24) depends on two assumptions: the variations of the partial specific volume of the protein during denaturation are negligible and the form factor $P_i(Q)$ (hence the shape) of the protein in a given state i does not vary in the temperature range of interest. Consequently, equation (23) will be valid only if the temperature range is narrow enough to ensure the constancy of the local electron density contrast $\rho(\vec{r}) - \rho_s$, and thereby the invariance of $P_i(Q)$. In the present work, equation (23) has been used over a temperature range of only 12 deg. C. Note that this can become a sensitive issue in the case of chemical denaturation, since ρ_s depends on the denaturing agent concentration in such a way that the form factor of the protein in a given state might undergo significant changes.

At low Q, the power expansion of the various terms of equation (23) yields the following relation:

$$R_g^2(T) = \sum_i f_i(T) R_i^2 \quad (24)$$

where R_i is the radius of gyration of species i and $R_g(T)$ is the experimental apparent radius of gyration at a given temperature T, determined from the experimental form factor $I(Q,T)/I(0,T)$. Equation (24) is valid under the same assumptions as equation (23).

Isoscattering point

When the normalised experimental scattering intensities measured along an unfolding process do not present a crossing-point common to all curves, the conclusion is easily reached that there must be at least one intermediate state. Indeed, for a strictly two-state transition from a native state, N, to an unfolded state, U, equation (23) shows that the total form factor $I(Q,T)/I(0,T)$ remains independent of T at the values of Q where the form factors of the native and unfolded states cross:

$$\frac{I(Q_0,T)}{I(0,T)} = (f_N(T) + f_U(T)) P_N(Q_0) = P_N(Q_0) \quad (25)$$

where Q_0 is a value of Q for which $P_N(Q) = P_U(Q)$.

However, the absence of isoscattering point proves the existence of intermediate species only when the normalized intensities $I(Q,T)/I(0,T)$ are considered. This might not be true with the experimental scattering intensities $I(Q,T)$. In the case of a two-state transition, the total scattering intensity becomes (from equation (22) and hypothesis of equation (23)):

$$I(Q,T) = c(T) K(T) (P_N(Q) + f_U(T)(P_U(Q) - P_N(Q))) \quad (26)$$

The scattering intensities of the individual states are necessarily measured at two different temperatures T_N and T_U, lower and higher than the temperature range of the transition, respectively. The equality observed at a crossing point Q_0 between these two curves must be written $K(T_N)P_N(Q_0) = K(T_U)P_U(Q_0)$, which, combined with equation (26) yields:

$$I(Q_0,T) = c(T) K(T) P_N(Q) \left(1 + f_U(T)\left(\frac{K(T_N)}{K(T_U)} - 1\right)\right) \quad (27)$$

Since $K(T)$ varies with temperature, so does $I(Q_0,T)$. Therefore, in spite of the two-state character of the transition, the scattering intensities $I(Q,T)$ do not present any isoscattering point (this is accentuated even more in the case of chemical unfolding, where the variation of the contrast factor, K, with the concentration of the denaturing agent is more pronounced).

The use of the form factors $I(Q,T)/I(0,T)$ is therefore an absolute prerequisite to relate the absence of isoscattering point to the existence of intermediate states in a meaningful way. In practice, it implies that $I(0,T)$ should be correctly derived from $I(Q,T)$, using the appropriate approximation for each curve (see Results).

We conclude this section by reminding the reader of the twofold hypothesis of the quasi-invariance of the specific volume and of the form factor of each species formulated in the previous paragraph. In particular, the absence of isoscattering point is meaningful only when a limited range of temperature (respectively of concentration of denaturing agent) is considered.

Singular value decomposition (SVD)

The SVD analysis is a very general method used to determine the number of independent species distinguishable by a given experimental technique, here SAXS, during a transition process, without prior knowledge of the patterns of the individual species. First applied to the analysis of SAXS patterns by Fowler *et al.*,[85,86] it has since been used several times in SAXS-monitored structural transitions and associations of biological macromolecules[87,88] and unfolding processes in particular.[38-46]

The set of N background-corrected experimental curves, of M channels each, are considered as the elements of a matrix $A(M \times N)$. The kth column of this matrix represents the scattering curve $I(Q,T_k)$ recorded at the kth temperature T_k and normalized by the intensity

at zero angle $I(0,T_k)$. This matrix can be decomposed as follows:

$$A(M \times N) = U(M \times N)w(N \times N)B(N \times N) \quad (28)$$

where the matrix U consists of N columns $\mathbf{u}_j(Q)$, which are the orthonormal basis vectors, the diagonal non-negative matrix w contains the singular values w_j of the decomposition, arranged in decreasing order, and B contains the normalized projections b_j^k of the kth experimental curve on the basis vector $\mathbf{u}_j$.

In order to increase the contribution of intensities in the medium Q range, which contains a wealth of relevant structural information, the SVD analysis was applied to the so-called Kratky profiles $Q^2I(Q)$. Following equation (28), each experimental pattern can be expressed as:

$$Q^2I(Q, T_k) = \sum_{j=1}^{N} w_j b_j^k \mathbf{u}_j(Q) \quad (29)$$

In most experimental situations, only the first L terms contribute significantly to the reconstruction of the experimental scattering profiles in equation (29). The w_j ($j > L$), which have very small values while the corresponding b_j^k display random fluctuations around zero as a function of k, only account for measuring noise. L represents the number of independent species distinguishable by SAXS during the transition or, in other words, the minimum number of thermodynamic states involved in the unfolding process.

The number L is determined by examining the evolution with temperature of the projections b_j^k of the experimental curves on the basis vector $\mathbf{u}_j$, the shape of the basis vectors $\mathbf{u}_j$, and checked by calculating the following χ^2 value, normalised by the experimental error bars $\sigma^k(Q)$:

$$\chi_p^2 = \frac{1}{(NM-1)} \sum_{k=1}^{N} \sum_{i=1}^{M} \left[\frac{(I(Q_i, T_k) - I_{\mathrm{calc,p}}(Q_i, T_k))^2}{\sigma^k(Q_i)^2} \right] \quad (30)$$

where $Q^2I_{\mathrm{calc,p}}(Q_i,T_k)$ is the best reconstruction of the kth experimental curve $Q^2I(Q_i,T_k)$ using the first p components in equation (29). L is the minimum value of p for which $\chi_p^2 \leqslant 1$.

Let us underline the fact that the basis vectors $\mathbf{u}_j(Q)$ are not the scattering curves of the L independent species or thermodynamic states. The determination of the latter curves, requires a second step using a parameterised model for the fractional concentrations of the various species, as described below.

Fitting procedure to determine the fractional concentration of the species present in solution at any temperature

When only three species are present in solution throughout the unfolding process, i.e. $L = 3$ (which is the case in the present work, as indicated in Results), a single intermediate is to be characterised. A procedure to retrieve its scattering pattern has been proposed by the group of Doniach.[42,44] The following draws heavily on their presentation, although a less constrained variant is proposed, designated first minimisation procedure. The three species correspond to three states of the protein, native (N), intermediate (I) and unfolded (U), and are characterised by their respective normalised scattering functions $I_N(Q)$, $I_I(Q)$ and $I_U(Q)$. In the particular case of our experiment, $I_N(Q)$ and $I_U(Q)$ are known and correspond to the experimental patterns at $T = T_1 = 61.6\,^\circ$C and $T = T_{10} = 76.8\,^\circ$C, respectively. On the contrary, $I_I(Q)$ is not known and its determination is expected to result from the following fitting procedure.

Each experimental Kratky scattering profile, $Q^2I(Q,T_k)$, can be decomposed uniquely on the SVD basis set:

$$Q^2I(Q, T_k) = w_1 b_1^k \mathbf{u}_1(Q) + w_2 b_2^k \mathbf{u}_2(Q) + w_3 b_3^k \mathbf{u}_3(Q) \quad (31)$$

where $w_1b_1^k$, $w_2b_2^k$ and $w_3b_3^k$ are the (known) projections deduced from the SVD analysis. But it can be written as the superposition of the profiles of the three species, weighted by their (unknown) respective fractional concentrations:

$$Q^2I(Q, T_k) = Q^2 f_N^k I_N(Q) + Q^2 f_I^k I_I(Q) + Q^2 f_U^k I_U(Q) \quad (32)$$

where $f_N^k + f_I^k + f_U^k = 1$. The scattering curves of the three independent species N, I and U may, in turn, be themselves decomposed on the SVD basis set:

$$\begin{aligned} Q^2I_N(Q) &= p_1^N \mathbf{u}_1(Q) + p_2^N \mathbf{u}_2(Q) + p_3^N \mathbf{u}_3(Q) \\ Q^2I_i(Q) &= p_1^I \mathbf{u}_1(Q) + p_2^I \mathbf{u}_2(Q) + p_3^I \mathbf{u}_3(Q) \\ Q^2I_U(Q) &= p_1^U \mathbf{u}_1(Q) + p_2^U \mathbf{u}_2(Q) + p_3^U \mathbf{u}_3(Q) \end{aligned} \quad (33)$$

where p_i^N and p_i^U are known but not the p_i^I.

Due to the linear independence of the SVD basis vectors, the decomposition of any scattering function on this basis set is unique. Therefore, the combination of equations (31) to (33) yields the relations that should be verified between the fractional concentrations f_N^k, f_I^k and f_U^k and the projections p_i^N, p_i^U and p_i^I on the one hand and the SVD-derived projections $w_jb_j^k$ on the other hand:

$$w_j b_j^k = f_N^k p_j^N + f_I^k p_j^I + f_U^k p_j^U \quad \text{for } j = 1, 2, 3 \quad (34)$$

Equation (34) is solved for the fractional concentrations f_N^k, f_I^k and f_U^k, and the projections p_1^I, p_2^I and p_3^I by minimising the following χ^2:

$$\chi^2 = \sum_{k=1}^{N} \sum_{j=1}^{L} [w_j (b_j^k - b_{j,\mathrm{calc}}^k)^2 / (\sigma_j^k)^2] \quad (35)$$

where $w_j b_{j,\mathrm{calc}}^k \equiv f_N^k p_j^N + f_I^k p_j^I + f_U^k p_j^U$ are the projection coefficients derived from equation (33).

The variances $(\sigma_j^k)^2$ are derived from the experimental variances $\sigma^2(Q,T_k)$ by the relation:

$$(\sigma_j^k)^2 = \sum_{i=1}^{M} Q_i^4 \sigma^2(Q_i, T_k) \mathbf{u}_j(Q_i)^2$$

To satisfy the summation condition $f_N^k + f_I^k + f_U^k = 1$, the species fractions are rewritten as:

$$f_N^k = \frac{1}{1+\exp(-a_k)+\exp(-b_k)}$$
$$f_I^k = \frac{\exp(-a_k)}{1+\exp(-a_k)+\exp(-b_k)} \quad (36)$$
$$f_U^k = \frac{\exp(-b_k)}{1+\exp(-a_k)+\exp(-b_k)}$$

The scattering pattern of the intermediate state I can then be reconstructed by using equation (33).

In the first minimisation procedure, no external constraints are put on a_k and b_k in equation (36). The summation condition is the only constraint that the fractional concentrations must satisfy. In this case, the solution to the minimisation problem is not unique. Rather, a class of solutions derivable by linear relationships can be found. This point is developed now. Let us consider a solution to the minimisation of equation (35), with $I_I(Q)$ being the pattern of the intermediate state and f_N^k, f_I^k, f_U^k the respective fractional concentrations. It is straightforward to show that the following pattern $J_I(Q)$, combined with the following fractional concentrations g_N^k, g_I^k, g_U^k yield the same right-hand side in equation (32) and therefore constitute another possible solution:

$$J_I(Q) = \alpha I_N(Q) + \beta I_U(Q) + (1-\alpha-\beta) I_I(Q)$$
$$g_N^k = f_N^k - f_I^k \frac{\alpha}{1-\alpha-\beta}$$
$$g_U^k = f_U^k - f_I^k \frac{\beta}{1-\alpha-\beta} \quad (37)$$
$$g_I^k = f_I^k - f_I^k \frac{1}{1-\alpha-\beta}$$

where α and β are scalars. For this solution to be minimally acceptable, the combination of α and β must be such that the values of g_N^k, g_I^k, g_U^k are between 0 and 1, and the patterns $J_I(Q)$ take only non-negative values. Given the latter restrictions, a set of possible solutions can be inferred from any solution provided by the minimisation procedure of equation (35). The uniqueness of the set of solutions derived in this way cannot be demonstrated formally. However, the different solutions provided by the minimisation algorithm with different initial guesses of the adjustable parameters all look very similar to representatives of the same set of linearly related solutions associated with one of them. Therefore, a single set of solutions is presented.

In the second minimisation procedure, originally reported by Chen *et al.*,[42] additional constraints are put upon a_k and b_k in equation (37), by considering thermodynamic relations between the three structural states. The free enthalpy differences between the native state and both the intermediate and the fully unfolded states can be written as a function of temperature:

$$\Delta G_{N\to I}(T) = \Delta H_{N\to I}(T) - T\Delta S_{N\to I}(T)$$
$$\Delta G_{N\to U}(T) = \Delta H_{N\to U}(T) - T\Delta S_{N\to U}(T) \quad (38)$$

This model assumes that the intermediate state is thermodynamically well defined. The fractional concentration of each species is then determined by replacing a_k and b_k in equation (36) by $\Delta G_{N\to I}(T_k)/RT_k$ and $\Delta G_{N\to U}(T_k)/RT_k$, respectively. If the variation with temperature of the respective differences in enthalpy and entropy is neglected, only four parameters, $\Delta H_{N\to I}$, $\Delta H_{N\to U}$, $\Delta S_{N\to I}$ and $\Delta S_{N\to U}$, remain adjustable in the minimisation procedure, whatever the number of experimental temperature values.

Acknowledgements

We are very grateful to E. Adjadj of the Institut Curie (Orsay) for sharing some of her results prior to publication. We thank P. Calmettes for discussions about the application of polymer theory to denatured proteins. We thank the technical staff of LURE-DCI and the operators of the storage ring. This work was supported by the CNRS through a grant of the PCV program.

References

1. Capaldi, A. P. & Radford, S. E. (1998). Kinetic studies of β-sheet protein folding. *Curr. Opin. Struct. Biol.* **8**, 86-92.
2. Grantcharova, V. P., Riddle, D. S. & Baker, D. (2000). Long-range order in the src SH3 folding transition state. *Proc. Natl Acad. Sci. USA,* **97**, 7084-7089.
3. Crane, J. C., Koepf, E. K., Kelly, J. W. & Gruebele, M. (2000). Mapping the transition state of the WW domain beta-sheet. *J. Mol. Biol.* **298**, 283-292.
4. Kortemme, T., Kelly, M. J., Kay, L. E., Forman-Kay, J. & Serrano, L. (2000). Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J. Mol. Biol.* **297**, 1217-1229.
5. Leeson, D. T., Gai, F., Rodriguez, H. M., Gregoret, L. M. & Dyer, R. B. (2000). Protein folding and unfolding on a complex energy landscape. *Proc. Natl Acad. Sci. USA,* **97**, 2527-2532.
6. Pappenberger, G., Saudan, C., Becker, M., Merbach, A. E. & Kiefhaber, T. (2000). Denaturant-induced movement of the transition state of protein folding revealed by high-pressure stopped-flow measurements. *Proc. Natl Acad. Sci. USA,* **97**, 17-22.
7. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I. & Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. *Nature Struct. Biol.* **6**, 1016-1024.
8. Martinez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Struct. Biol.* **6**, 1010-1016.
9. Cota, E., Hamill, S. J., Fowler, S. B. & Clarke, J. (2000). Two proteins with the same structure respond very differently to mutation: the role of plasticity in protein stability. *J. Mol. Biol.* **302**, 713-725.
10. Hamill, S. J., Steward, A. & Clarke, J. (2000). The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *J. Mol. Biol.* **297**, 165-178.
11. Hamill, S. J., Cota, E., Chothia, C. & Clarke, J. (2000). Conservation of folding and stability within a protein family: the tyrosine corner as an evolutionary cul-de-sac. *J. Mol. Biol.* **295**, 641-649.
12. Cota, E. & Clarke, J. (2000). Folding of beta-sandwich proteins: three-state transition of a fibronectin type III module. *Protein Sci.* **9**, 112-120.

13. Clarke, J., Cota, E., Fowler, S. B. & Hamill, S. J. (1999). Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. *Struct. Fold. Des.* **7**, 1145-1153.
14. Clarke, J., Hamill, S. J. & Johnson, C. M. (1997). Folding and stability of a fibronectin type III domain of human tenascin. *J. Mol. Biol.* **270**, 771-778.
15. Adjadj, E., Quiniou, E., Mispelter, J., Favaudon, V. & Lhoste, J. M. (1992). The seven-stranded beta-barrel structure of apo-neocarzinostatin as compared to the immunoglobulin domain. *Biochimie,* **74**, 853-858.
16. Adjadj, E., Quiniou, E., Mispelter, J., Favaudon, V. & Lhoste, J. M. (1992). Three-dimensional solution structure of apo-neocarzinostatin from Streptomyces carzinostaticus determined by NMR spectroscopy. *Eur. J. Biochem.* **203**, 505-511.
17. Gibson, B. W., Herlihy, W. C., Samy, T. S., Hahm, K. S., Maeda, H., Meienhofer, J. & Biemann, K. (1984). A revised primary structure for neocarzinostatin based on fast atom bombardment and gas chromatographic-mass spectrometry. *J. Biol. Chem.* **259**, 10801-10806.
18. Napier, M. A., Holmquist, B., Strydom, D. J. & Goldberg, I. H. (1979). Neocarzinostatin: spectral characterization and separation of a non-protein chromophore. *Biochem. Biophys. Res. Commun.* **89**, 635-642.
19. Favaudon, V., Charnas, R. L. & Goldberg, I. H. (1985). Poly(deoxyadenylic-deoxythymidylic acid) damage by radiolytically activated neocarzinostatin. *Biochemistry,* **24**, 250-259.
20. Povirk, L. F. & Goldberg, I. H. (1980). Binding of the nonprotein chromophore of neocarzinostatin to deoxyribonucleic acid. *Biochemistry,* **19**, 4773-4780.
21. Russo, D., Durand, D., Calmettes, P. & Desmadril, M. (2001). Characterization of the denatured states distribution of neocarzinostatin by small-angle neutron scattering and differential scanning calorimetry. *Biochemistry,* **40**, 3958-3966.
22. Gualfetti, P. J., Iwakura, M., Lee, J. C., Kihara, H., Bilsel, O., Zitzewitz, J. A. & Matthews, C. R. (1999). Apparent radii of the native, stable intermediates and unfolded conformers of the alpha-subunit of tryptophan synthase from *E. coli*, a TIM barrel protein. *Biochemistry,* **38**, 13367-13378.
23. Kataoka, M., Kuwajima, K., Tokunaga, F. & Goto, Y. (1997). Structural characterization of the molten globule of alpha-lactalbumin by solution X-ray scattering. *Protein Sci.* **6**, 422-430.
24. Fang, X., Littrell, K., Yang, X., Henderson, S. J., Siefert, S., Thiyagarajan, P., Pan, T. & Sosnick, T. R. (2000). Mg^{2+}-dependent compaction and folding of yeast tRNA(Phe) and the catalytic domain of the B. subtilis RNase P RNA determined by small-angle X-ray scattering. *Biochemistry,* **39**, 11107-11113.
25. Russell, R., Millett, I. S., Doniach, S. & Herschlag, D. (2000). Small angle X-ray scattering reveals a compact intermediate in RNA folding. *Nature Struct. Biol.* **7**, 367-370.
26. Konno, T., Kamatari, Y. O., Tanaka, N., Kamikubo, H., Dobson, C. M. & Nagayama, K. (2000). A partially unfolded structure of the alkaline-denatured state of pepsin and its implication for stability of the zymogen-derived protein. *Biochemistry,* **39**, 4182-4190.
27. Receveur, V., Garcia, P., Durand, D., Vachette, P. & Desmadril, M. (2000). Role of hydrophobic interactions in yeast phosphoglycerate kinase stability. *Proteins: Struct. Funct. Genet.* **38**, 226-238.
28. Gast, K., Damaschun, H., Misselwitz, R., Muller-Frohne, M., Zirwer, D. & Damaschun, G. (1994). Compactness of protein molten globules: temperature-induced structural changes of the apomyoglobin folding intermediate. *Eur. Biophys. J.* **23**, 297-305.
29. Flanagan, J. M., Kataoka, M., Fujisawa, T. & Engelman, D. M. (1993). Mutations can cause large changes in the conformation of a denatured protein. *Biochemistry,* **32**, 10359-10370.
30. Uversky, V. N., Karnoup, A. S., Segel, D. J., Seshadri, S., Doniach, S. & Fink, A. L. (1998). Anion-induced folding of Staphylococcal nuclease: characterization of multiple equilibrium partially folded intermediates. *J. Mol. Biol.* **278**, 879-894.
31. Kamatari, Y. O., Ohji, S., Konno, T., Seki, Y., Soda, K., Kataoka, M. & Akasaka, K. (1999). The compact and expanded denatured conformations of apomyoglobin in the methanol-water solvent. *Protein Sci.* **8**, 873-882.
32. Bonneté, F., Madern, D. & Zaccai, G. (1994). Stability against denaturation mechanisms in halophilic malate dehydrogenase ''adapt'' to solvent conditions. *J. Mol. Biol.* **244**, 436-447.
33. Panick, G., Vidugiris, G. J., Malessa, R., Rapp, G., Winter, R. & Royer, C. A. (1999). Exploring the temperature-pressure phase diagram of staphylococcal nuclease. *Biochemistry,* **38**, 4157-4164.
34. Pollack, L., Tate, M. W., Darnton, N. C., Knight, J. B., Gruner, S. M., Eaton, W. A. & Austin, R. H. (1999). Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle X-ray scattering. *Proc. Natl Acad. Sci. USA,* **96**, 10115-10117.
35. Plaxco, K. W., Millett, I. S., Segel, D. J., Doniach, S. & Baker, D. (1999). Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nature Struct. Biol.* **6**, 554-556.
36. Eliezer, D., Jennings, P. A., Wright, P. E., Doniach, S., Hodgson, K. O. & Tsuruta, H. (1995). The radius of gyration of an apomyoglobin folding intermediate. *Science,* **270**, 487-488.
37. Arai, M., Ikura, T., Semisotnov, G. V., Kihara, H., Amemiya, Y. & Kuwajima, K. (1998). Kinetic refolding of beta-lactoglobulin. studies by synchrotron X-ray scattering, and circular dichroism, absorption and fluorescence spectroscopy. *J. Mol. Biol.* **275**, 149-162.
38. Chen, L., Wildegger, G., Kiefhaber, T., Hodgson, K. O. & Doniach, S. (1998). Kinetics of lysozyme refolding: structural characterization of a non-specifically collapsed state using time-resolved X-ray scattering. *J. Mol. Biol.* **276**, 225-237.
39. Segel, D. J., Eliezer, D., Uversky, V., Fink, A. L., Hodgson, K. O. & Doniach, S. (1999). Transient dimer in the refolding kinetics of cytochrome c characterized by small-angle X-ray scattering. *Biochemistry,* **38**, 15352-15359.
40. Segel, D. J., Bachmann, A., Hofrichter, J., Hodgson, K. O., Doniach, S. & Kiefhaber, T. (1999). Characterization of transient intermediates in lysozyme folding with time-resolved small-angle X-ray scattering. *J. Mol. Biol.* **288**, 489-499.
41. Hoshino, M., Hagihara, Y., Hamada, D., Kataoka, M. & Goto, Y. (1997). Trifluoroethanol-induced conformational transition of hen egg-white lysozyme studied by small-angle X-ray scattering. *FEBS Letters,* **416**, 72-76.

42. Chen, L., Hodgson, K. O. & Doniach, S. (1996). A lysozyme folding intermediate revealed by solution X-ray scattering. *J. Mol. Biol.* **261**, 658-671.
43. Hagihara, Y., Hoshino, M., Hamada, D., Kataoka, M. & Goto, Y. (1998). Chain-like conformation of heat-denatured ribonuclease A and cytochrome *c* as evidenced by solution X-ray scattering. *Fold. Des.* **3**, 195-201.
44. Segel, D. J., Fink, A. L., Hodgson, K. O. & Doniach, S. (1998). Protein denaturation: a small-angle X-ray scattering study of the ensemble of unfolded states of cytochrome c. *Biochemistry*, **37**, 12443-12451.
45. Koide, S., Bu, Z., Risal, D., Pham, T. N., Nakagawa, T., Tamura, A. & Engelman, D. M. (1999). Multistep denaturation of *Borrelia burgdorferi* OspA, a protein containing a single-layer beta-sheet. *Biochemistry*, **38**, 4757-4767.
46. Kojima, M., Tanokura, M., Maeda, M., Kimura, K., Amemiya, Y., Kihara, H. & Takahashi, K. (2000). pH-Dependent unfolding of aspergillopepsin II studied by small-angle X-ray scattering. *Biochemistry*, **39**, 1364-1372.
47. des, Cloiseaux J. & Jannink, G. (1990). *Polymers in Solution: Their modelling and their structure*, Clarendon Press, Oxford.
48. Douglas, J. F. & Freed, K. F. (1984). Penetration function and second virial coefficient for linear and regular star polymers. *Macromolecules*, **17**, 1854-1870.
49. Bonneté, F., Finet, S. & Tardieu, A. (1999). Second virial coefficient: variations with lysozyme crystallisation conditions. *J. Crystal Growth*, **196**, 403-414.
50. Svergun, D. I., Barberato, C. & Koch, M. H. J. (1995). CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallog.* **28**, 768-773.
51. Svergun, D. I., Richard, S., Koch, M. H., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). Protein hydration in solution: experimental observation by X-ray and neutron scattering. *Proc. Natl Acad. Sci. USA*, **95**, 2267-2272.
52. Russo, D., Durand, D., Desmadril, M. & Calmettes, P. (2000). Study of thermally and chemically unfolded conformations of a small β-protein by means of small-angle neutron scattering. *Physica B*, **276-278**, 520-521.
53. Gripon, C., Legrand, L., Rosenman, I., Vidal, O., Robert, M. C. & Boué, F. (1997). Lysozyme-lysozyme interactions in under- and super-saturated solutions: a simple relation between the virial coefficients in H_2O and ^2H_2O. *J. Crystal Growth*, **178**, 575-584.
54. Flory, P. (1971). *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, NY.
55. Kharakoz, D. P. (1997). Partial volumes and compressibilities of extended polypeptide chains in aqueous solution: additivity scheme and implication of protein unfolding at normal and high pressure. *Biochemistry*, **36**, 10276-10285.
56. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81-R91.
57. Kataoka, M., Head, J. F., Vorherr, T., Krebs, J. & Carafoli, E. (1991). Small-angle X-ray scattering study of calmodulin bound to two peptides corresponding to parts of the calmodulin-binding domain of the plasma membrane Ca^{2+} pump. *Biochemistry*, **30**, 6247-6251.
58. Cantor, C. R. & Schimmel, P. R. (1980). Part III: the behavior of biological macromolecules. In *Biophysical Chemistry*, p. 1011, W. H. Freeman and Company, New York.
59. Yi, Q., Scalley-Kim, M. L., Alm, E. J. & Baker, D. (2000). NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Biol.* **299**, 1341-1351.
60. Mok, Y. K., Alonso, L. G., Lima, L. M., Bycroft, M. & de Prat-Gay, G. (2000). Folding of a dimeric beta-barrel: residual structure in the urea denatured state of the human papillomavirus E2 DNA binding domain. *Protein Sci.* **9**, 799-811.
61. Wong, K. B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M., Fersht, A. R. & Daggett, V. (2000). Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol.* **296**, 1257-1282.
62. Sari, N., Alexander, P., Bryan, P. N. & Orban, J. (2000). Structure and dynamics of an acid-denatured protein G mutant. *Biochemistry*, **39**, 965-977.
63. Miller, W. G. & Goebel, C. V. (1968). Dimensions of protein random coils. *Biochemistry*, **7**, 3925.
64. Rowe, G. & Lopez Pineiro, A. (1990). Influence of the solvent on the conformational-dependent properties of random-coil polypeptides. I. The mean-square of the end-to-end distance and of the dipole moment. *Biophys. Chem.* **36**, 57-64.
65. Kirste, R. G. & Oberthür, R. C. (1982). Synthetic polymers in solution. In *Small Angle X-ray Scattering* (Glatter, O. & Kratky, O., eds), pp. 387-431, Academic Press, London, New York.
66. Chalikian, T. V. & Breslauer, K. J. (1996). On volume changes accompanying conformational transitions in biopolymers. *Biopolymers*, **39**, 619-626.
67. Filfil, R. & Chaliikian, T. V. (2000). Volumetric and spectroscopic characterizations of the native and acid-induced denatured states of staphylcoccal nuclease. *J. Mol. Biol.* **299**, 827-842.
68. Onuchic, J. N., Nymeyer, H., Garcia, A. E., Chahine, J. & Socci, N. D. (2000). The energy landscape theory of protein folding: insights into folding mechanisms and scenarios. *Advan. Protein Chem.* **53**, 87-152.
69. Griko, Y. V. (1999). Denaturation versus unfolding: energetic aspects of residual structure in denatured alpha-lactalbumin. *J. Protein Chem.* **18**, 361-369.
70. Griko, Y. V., Freire, E. & Privalov, P. L. (1994). Energetics of the alpha-lactalbumin states: a calorimetric and statistical thermodynamic study. *Biochemistry*, **33**, 1889-1899.
71. Hirai, M., Arai, S., Iwase, H. & Takizawa, T. (1998). Small-angle X-ray scattering and calorimetric studies of thermal conformational change of lysozyme depending on pH. *J. Phys. Chem. ser. B*, **102**, 1308-1313.
72. Heyd, B., Lerat, G., Adjadj, E., Minard, P. & Desmadril, M. (2000). Reinvestigation of the proteolytic activity of neocarzinostatin. *J. Bacteriol.* **182**, 1812-1818.
73. Dubuisson, J. M., Decamps, T. & Vachette, P. (1997). Improved signal-to-background ratio in small-angle X-ray scattering experiments with synchrotron radiation using an evacuated cell for solutions. *J. Appl. Crystallog.* **30**, 49-54.
74. Calmettes, P., Durand, D., Desmadril, M., Minard, P., Receveur, V. & Smith, J. C. (1994). How random is a highly denatured protein? *Biophys. Chem.* **53**, 105-114.
75. Guinier, A. & Fournet, G. (1955). *Small Angle Scattering of X-Rays*, Wiley, New York.
76. Debye, P. (1947). *J. Phys. Colloid. Chem.* **51**, 18-32.

77. Svergun, D., Semenyuk, ?. & Feigin, L. A. (1988). Small-angle-scattering-data treatment by the regularization method. *Acta Crystallog. sect. A*, **44**, 244-250.
78. Kratky, O. & Porod, G. (1949). Röntgenuntersuchung Gelöster Fadenmoleküle. *Rec. Trav. Chim. Pays-Bas*, **68**, 1106-1122.
79. Sharp, P. & Bloomfield, V. A. (1968). Light scattering from wormlike chains with excluded volume effects. *Biopolymers*, **6**, 1201-1211.
80. des Cloiseaux, J. (1973). Form factor of an infinite Kratky-Porod chain. *Macromolecules*, **6**, 403-407.
81. Rawiso, M., Duplessix, R. & Picot, C. (1987). Scattering function of polystyrene. *Macromolecules*, **20**, 630-648.
82. Pedersen, J. S. & Schurtenberger, P. (1996). Scattering functions of semiflexible polymers with and without excluded volume effects. *Macromolecules*, **29**, 7602-7612.
83. Porod, G. (1951). *Kolloid-Z.* **124**, 83-114.
84. Le Guillou, J. C. & Zinn-Justin, J. (1977). Critical exponents for the n-vector model in three dimensions from field theory. *Phys. Rev. Letters*, **39**, 95-98.
85. Fowler, A. G., Foote, A. M., Moody, M. F., Vachette, P., Provencher, S. W., Gabriel, A., Bordas, J. & Koch, M. H. J. (1983). Stopped-flow solution scattering using synchrotron radiation: apparatus, data collection and data analysis. *J. Biochem. Biophys. Methods*, **7**, 317-329.
86. Provencher, S. W. & Glöckner, J. (1983). Analysis of components present in kinetics (or titration) curves. *J. Biochem. Biophys. Methods*, **7**, 331-334.
87. Fetler, L., Tauc, P., Hervé, G., Moody, M. F. & Vachette, P. (1995). X-ray scattering titration of the structure transition of aspartate transcarbamylase with a bisubstrate analogue. Influence of nucleotide effectors. *J. Mol. Biol.* **251**, 243-255.
88. Bilgin, N., Ehrenberg, M., Ebel, C., Zaccai, G., Sayers, Z., Koch, M. H. J., Svergun, D. I., Barberato, C., Volkov, V., Nissen, P. & Nyborg, J. (1998). Solution structure of the ternary complex between aminoacyl-tRNA, elongation factor Tu, and guanosine triphosphate. *Biochemistry*, **37**, 8163-8172.

Appendix

Critical Confrontation of Three Methods of Calculation of the Apparent Radius of Gyration of a Mixture of Native and Unfolded Protein

To evaluate accurately the average radius of gyration of the protein as a function of temperature during the unfolding transition, the most appropriate analysis has to be chosen. Three approaches can be considered; Guinier's law, Debye's formula and the calculation of the distance distribution function $P(r)$ (see Materials and Methods of the main text). Indeed, the simultaneous presence of different states of the protein in the solution (compact, unfolded and possibly partially unfolded) makes this question not a straightforward one. To clarify the issue, we use all three methods to calculate the radius of gyration of synthetic scattering patterns $I_{sy}(Q)$, which are linear combinations (coefficients $1-f_U$, f_U) of the native ($R_g = R_N$) and the unfolded ($R_g = R_U$) protein scattering curves, thereby simulating a two-state transition between a globular and an extended state. The comparison of the obtained values with the true value given by $R_{sy} = \sqrt{(1-f_U)R_N^2 + f_U R_U^2}$ then provides us with a direct assessment of the way each method performs.

In the absence of interactions, the intensity is given by the following expression:

$$I_{sy}(Q) = (1-f_U)I_N(Q) + F_U I_U(Q) \qquad \text{(A1)}$$

where f_U is the fraction of unfolded protein. $I_N(Q)$ is the experimental intensity scattered by the protein in its native conformation, $I_U(Q)$ is the intensity scattered by the unfolded protein. The synthetic intensity $I_{sy}(Q)$ is calculated for different values of f_U. An apparent radius of gyration is then extracted from the curves using either the Guinier or the Debye approximation in the range $0.02\ \text{Å}^{-1} < Q < 0.055\ \text{Å}^{-1}$. This angular range is deemed reasonable, since, for a value of 26 Å, $Q_{max} = 0.055\ \text{Å}^{-1}$ corresponds to $Q_{max}R_g < 1.4$. The two estimates $R_{Guinier}$ and R_{Debye} of the average radius of gyration and the true value of this parameter given by $R_{sy} = \sqrt{(1-f_U)R_N^2 + f_U R_U^2}$ with $R_N = 14.0$ Å and $R_U = 26.1$ Å, have been determined for different values of f_U. Finally, the $P(r)$ function has been calculated for each curve and the corresponding apparent radius of gyration, R_{Pr}, extracted. The results, which are presented in Figure A1, call for several comments.

(i) R_{Debye} yields a slight underestimate of the ''true'' value, the difference being largest at the centre of the transition, whilst $R_{Guinier}$ consistently underestimates the value of the average radius of

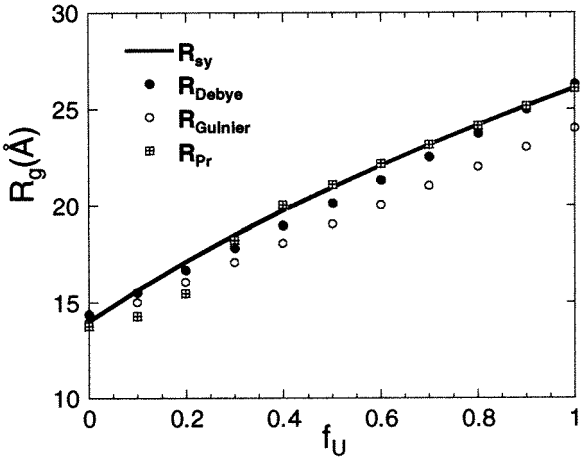

Figure A1. Determination of the apparent radius of gyration of a mixture of native and fully unfolded protein using three methods. Open circles, Guinier's law; filled circles, Debye's formula; crosses in square, distance distribution function. Continuous line, ''true'' value $R_{sy} = \sqrt{(1-f_U)R_N^2 + f_U R_U^2}$ using the Guinier determination for the native protein $R_N = 14$ Å and the Debye value for the unfolded protein $R_U = 26.1$ Å (see the text for details).

gyration, the discrepancy increasing with the proportion of unfolded protein present in the solution. Of course, Guinier's law still holds true for an extended chain, but its range of validity is restricted to such small-angles as being essentially out of experimental reach. This is due to the existence of very large distances between two points of the unfolded protein which makes the Q^4 term in the power expansion of $I(Q)$ contribute significantly even at low Q value, thereby reducing the range of validity of the Guinier approximation. Conversely, applying Debye's formula over a range larger than $Q_{max}R_g < 1.4$ would lead to an increasing discrepancy between the resulting estimate of the radius of gyration and its actual value.

It is interesting that R_{Pr} is practically identical with the true value, with the exception of the first two points, at the onset of the transition, where its value strongly underestimates the actual value of R_g. The marked discrepancy observed for $0.1 \leqslant f_U \leqslant 0.2$ is due to the presence of large, unfolded particles in small amount. The distribution of these large vectors involves the determination of a number of Fourier coefficients, which is made difficult by the weak contribution of these vectors to the scattered intensity. Indeed, if the value of D_{max} found for the $P(r)$ of the unfolded protein ($f_U = 1$), $D_{max} = 106$ Å, is used in the calculation of the $P(r)$ of the curves with $0.1 \leqslant f_U \leqslant 0.2$, as it should be, the shape of the $P(r)$ display strong fluctuations at the higher values of r. The effective values of D_{max} below which no such fluctuations appear are actually 50 Å and 60 Å for $f_U = 0.1$ and $f_U = 0.2$, respectively. Even for $f_U = 0.3$, where R_{Pr} is very close to the true value of R_g, the value of the apparent D_{max} is found to be 91 Å, not 106 Å. One should therefore be very careful with the significance attributed to the ''experimental'' values of D_{max}.

(ii) The variation of the apparent R_g with f_U appears to be essentially linear when using either Guinier's law or Debye's formula, while it is well known that the average value of R_g is the quadratic mean of the R_g of the two mixed species weighted by their respective fractional concentration. This simply proves that none of the two laws is adequate to provide the exact value of the true R_g (i.e. the quadratic mean), though for different reasons. For Guinier's law, the problem is not theoretical (the linear combination of two Gaussian curves will yield the proper average) but experimental (the region of validity of the approximation when unfolded protein is present is essentially hidden behind the beam-stop, and thus not accessible). For Debye's formula, the difficulty is mathematical, since the linear combination of two ''Debye profiles'' does not directly yield the appropriate combination of radii of gyration. Finally, R_{Pr} does follow the proper evolution with f_U not only theoretically, as recently underlined by Fang *et al.*,[24] but also practically, since its value is very close to the actual value with the exception of the low values of f_U already mentioned.

Some practical conclusions can thus be drawn from our simple calculations. The derivation of R_g from the $P(r)$ is to be preferred to extract the average radius of gyration from a mixture of native and unfolded proteins, but it is unreliable at the onset of the transition. In this range, it is best supplemented by Debye's formula, which consistently yields better results than Guinier's law, provided the fitting range is restricted enough ($Q_{max}R_g < 1.4$). The maximum value between R_{Pr} and R_{Debye} should be chosen in any case.

Let us emphasise again that in the case of a two-state transition, the value of the apparent R_g derived from either Guinier's or Debye's formula presents an essentially linear variation with the fraction of unfolded protein instead of the expected parabolic dependence. This variation must thus be considered with extreme caution, as well as its use to derive thermodynamic parameters. If intermediate species are detected, the variation of the apparent R_g with f_U will be even more difficult to relate to the actual transition.

Edited by M. F. Moody

(Received 9 November 2000; received in revised form 14 February 2001; accepted 13 March 2001)